

**Structural investigation of the GAF
domain protein BPSL2418 from
*Burkholderia pseudomallei***

Mona Ghonaim Alharbi

M.Sc, B.Sc. (Hons) Biology

**A thesis submitted in partial fulfilment of the requirements for the
degree of Doctor of Philosophy**

**Department of Molecular Biology and Biotechnology,
University of Sheffield**

September 2014

Abstract

A new family of methionine-sulfoxide reductase (Msr) was recently discovered, and was named free methionine sulfoxide reductase (fRMsr). This family includes enzymes with a reductase activity toward the free *R* isomer of a methionine sulfoxide substrate. The fRMsr have a GAF domain topology, a domain, which was previously identified as having in some cases a cyclic nucleotide phosphodiesterase activity. The classification of fRMsr as GAF domains revealed a new function can be added to the GAF domain family. Interestingly the four members identified in the fRMsr family share the GAF domain structure and the presence of three conserved cysteines in the active site with free *R* methionine sulfoxide substrate specificity. This thesis presents the crystal structures of reduced, free Met-SO substrate-bound and MES-bound forms of a new fRMsr from *Burkholderia pseudomallei* (BPSL2418). BPSL2418 was cloned, overexpressed and purified to enable protein crystallization. The crystallization trials for reduced, Met-SO-bound and MES-bound forms of BPSL2418 were prepared and reasonable crystals of each form were produced. The crystal structures of BPSL2418_{MES}, BPSL2418_{Met-SO} and BPSL2418_{Reduced} were solved at 1.18, 1.4 and 2.0Å, respectively by molecular replacement. The BPSL2418_{MES} crystal belongs to space group P 2₁ 2₁ 2₁ while BPSL2418_{Met-SO} and BPSL2418_{Reduced} crystals belong to space group P 1 2₁ 1. All three forms share the GAF domain structure of six antiparallel β-strands and four α-helices with connecting loops. The antiparallel β-strands (β1, β2, β5 and β6) are located in the center of the BPSL2418 structure flanked on one side by a three α-helices (α1, α2 and α4) and on the other side by a (loop1, β3, loop2, α3, β4 loop4) unit where loop4 forms a capping flap and covers the active site. The structural comparison of the three forms of BPSL2418 indicates that the catalytically important cysteine is CYS109, where the resolving cysteine is CYS75, which forms a disulfide bond with CYS109. They also suggest that the third conserved cysteine in the active site, CYS85, which is located in α3, is a non-essential cysteine for the catalytic function but it may play a role in the binding of the substrate. The structural comparison of the three forms reveals that conformational changes appear in the active site particularly involving loop4 and CYS109 during catalysis. The 3D structure of BPSL2418 shows strong structure similarity to fRMsr enzymes, which further suggests that BPSL2418 acts as a free Met-*R*-SO reductase and shares the catalytic mechanism of fRMsr family.

Acknowledgement

First of all, I praise the almighty God for giving me the strength and guidance that entailed me to fulfill the requirement to present this work.

My sincerely and immense gratitude to my previous supervisor Professor Peter Artymiuk for his encouragement, inspiration, guidance and invaluable and friendly advice from the beginning of my PhD until we sadly lost him on February 2014.

I would like to express my great appreciation and deeply thanks to my supervisor Dr. John Rafferty for his support and guidance at the final stage of this work, which made this thesis possible. I am also grateful to Professor David Rice for advice provided to me whenever I requested. I am indebted to Dr. Patrick Baker for his valuable help and guidance particularly in crystallization and data collection. I am also deeply indebted to Dr. Svetlana Sedelnikova for her tireless manipulation of protein purification and her general advice in laboratory work, without their support obtaining results won't be possible. A special thanks to Fiona Rodger for her kindness and the wonderful lab organization, which offer a great place to work in. Many thanks to all the other members of the X-ray crystallography group particularly Dr. Claudine Bisson for her help and support. Also my great appreciation to Professor Albert Ong, Dr. Frank Xu from the Medical school and Dr. Svetomir Tzokov for helping in electron microscope using.

I would like also to express my very special thanks to all my friends and colleague Reem, Magdah, Feras, Abdulhamid and Elbandari for their support and encouragement especially during difficult time.

I am truly indebted to my parents, Ghonaim and Salwa for all unconditional love, support, care and sacrifices. There is no words can explain my gratitude towards them. Many thanks also for my wonderful sisters and brother. Very heartfelt thanks to my other me, my soul mate and husband Rashad for his love, great support, sacrifices and sharing my good and bad times during my PhD journey. Finally, to Khalid, Tala and Ahmed my heart and my little support system, I love you.

This thesis is dedicated to my beloved parents, husband and children.

1.1.5.2 Capsular polysaccharides	35
1.1.5.3 Lipopolysaccharide	36
1.1.5.4 Flagella	36
1.1.5.5 Pili	37
1.1.5.6 Biofilm formation.....	37
1.1.5.7 Antibiotic resistance.....	37
1.1.5.8 Secretion systems	38
1.1.5.9 The intracellular lifestyle of <i>B. pseudomallei</i>	40
1.2 Methionine sulfoxide reductases, members of GAF domain family.....	41
1.2.1 The GAF family	41
1.2.1.1 GAF family history	41
1.2.1.2 GAF domain structure.....	42
1.2.1.3 GAF domain functions.....	44
1.2.2 Methionine sulfoxide reductases.....	46
1.2.2.1 Reactive oxygen species (ROS) and oxidative stress	46
1.2.2.2 Oxidation of methionine	47
1.2.2.3 Methionine sulfoxide reductase (Msr) family.....	48
1.2.2.4 Role of methionine sulfoxide reductases in bacterial pathogenicity	51
1.2.3 Examples of methionine sulfoxide reductases containing GAF domain	51
1.2.4 A new free methionine-(<i>R</i>)-sulfoxide reductase from <i>Burkholderia pseudomallei</i> (BPSL2418) belongs to the GAF domain family	58
1.3 Aims of the projects	64
Second project: The C-terminal domains of Polycystin1 (CT1) and Polycystin2 (CT2).....	65
1.4 Polycystin1 and Polycystin2	65
1.4.1 Autosomal Dominant Polycystic Kidney Disease (ADPKD):.....	65

1.4.2 Polycystin1	66
1.4.2.1 Domain structure of Polycystin1	66
1.4.2.2 Cellular and subcellular localization of Polycystin1	69
1.4.2.3 Function of Polycystin1	69
1.4.3 Polycystin2	71
1.4.3.1 Domain structure of Polycystin2	71
1.4.3.2 Cellular and subcellular localization of Polycystin2	74
1.4.3.3 Function of Polycystin2	74
1.4.4 Polycystin1 and polycystin2 interaction	74
1.5 Aims of the project	77
Chapter 2: Materials and Methods.....	78
2.1 Bacterial strains, DNA, plasmids and oligonucleotides	78
2.2 Materials	79
2.3 Microbiological Methods	79
2.3.1 Growth media	79
2.4 Polymerase chain reaction (PCR).....	81
2.5 Agarose gel electrophoresis	82
2.6 Purification of PCR products by gel extraction:	83
2.7 Plasmid purification.....	83
2.8 Cloning of expression construct.....	84
2.8.1 Cloning with pETBlue-1 vector	84
2.8.2 Ligation and transformation for pETBlue-1 cloning	86
2.8.3 Confirmation of cloning results	86
2.9 Transformation	87
2.10 Protein expression	87
2.10.1 Regulation of protein expression in the pET expression system	88

2.10.2 Protein overexpression protocol.....	88
2.10.2.1 Small-scale overexpression trials.....	88
2.10.2.2 Large-scale overexpression trials.....	89
2.11 Sodium dodecyl sulfate Polyacrylamide gel electrophoresis (SDS-PAGE)	90
2.12 Solubilisation and refolding of proteins	92
2.12.1 Inclusion bodies preparation	92
2.12.2 Solubilisation and refolding of inclusion bodies.....	93
2.12.2.1 Refolding CT1.....	93
2.13 Protein purification	93
2.13.1 Ion exchange chromatography	94
2.13.2 Hydrophobic interaction chromatography	94
2.13.3 Gel filtration.....	95
2.13.4 Nickel Nitrilotriacetic acid (Ni-NTA) Sepharose	97
2.14 Protein concentration.....	97
2.15 Protein crystallization to structure identification	98
2.15.1 Producing crystals	98
2.15.2 Cryoprotection of crystals	98
2.15.3 Data collection and processing.....	100
2.15.4 Structure determination.....	100
2.16 Electron Microscopy	101
Chapter 3: Theory of X-Ray Crystallography.....	102
3.1 The use of X-rays to produce images of protein molecules	102
3.2 The features of crystals.....	103
3.3 Producing protein crystals	104
3.3.1 Crystallization techniques	106

3.3.1.1 Hanging drop technique	106
3.3.1.2 Sitting drop technique	108
3.4 Crystal mounting.....	108
3.4.1 Cryo-cooling protein crystals.....	108
3.4.2 The requirement for a cryo-protectant	110
3.5 Collecting X-ray data.....	110
3.6 X-ray sources	111
3.6.1 Rotating copper anode tubes	111
3.6.2 Synchrotron.....	113
3.7 X-ray Detectors.....	113
3.7.1 Image plate detectors.....	113
3.7.2 Charged-coupled device (CCD) detectors	115
3.8 The principles of X-ray diffraction.....	116
3.8.1 Diffraction from crystals.....	116
3.8.2 Bragg's Law	116
3.9 From diffraction to electron density.....	118
3.10 Processing data from diffraction images.....	119
3.10.1 Data indexing and integration	119
3.10.2 Data merging and scaling.....	119
3.11 Structural determination.....	119
3.11.1 initial phasing.....	119
3.11.1.1 Molecular replacement (MR).....	120
3.12 Structure re-building and refinement	121
3.13 Structure validation	121
Chapter 4: BPSL2418 Cloning, Expression, Purification and Crystallization	123

4.1 Cloning of BPSL2418.....	123
4.1.1 Amplification of the <i>bpsl2418</i> gene.....	123
4.1.2 Ligation of <i>bpsl2418</i> gene into pETBlue-1 vector and transformation into Novablue competent <i>E. coli</i> cells.....	123
4.1.3 Confirmation of the presence and orientation of the insert gene in the pETBlue-1 plasmid	125
4.2 Transformation of Tuner (DE3)pLacI competent cell with pETBlue-1 <i>bpsl2418</i> recombinants.....	125
4.3 BPSL2418 overexpression	125
4.4 Purification of BPSL2418	128
4.5 Crystallization trials of BPSL2418	136
4.5.1 Crystallization of the BPSL2418 MES-bound form (BPSL2418 _{MES}) .	136
4.6 Crystallization of BPSL2418 unbound, reduced and substrate-bound forms.....	138
4.6.1 Crystallization of BPSL2418 unbound form (BPSL2418 _{unbound})	138
4.6.2 Crystallization of BPSL2418 substrate-bound form (BPSL2418 _{Met-SO})	141
4.6.3 Crystallization of BPSL2418 reduced form (BPSL2418 _{Reduced})	141
4.7 Cryo-protection for BPSL2418 forms	144
Chapter 5: Structure determination of BPSL2418	145
5.1 Experimental structure determination of BPSL2418_{MES} by molecular replacement using PDB entry: 3ksf.....	145
5.1.1 Space group determination and indexing.....	145
5.1.2 Data collection and processing.....	146
5.1.3 Matthews Coefficient (<i>V_m</i>) calculation	146
5.1.4 Structure determination of BPSL2418 _{MES}	146
5.1.5 Model building and refinement.....	150

5.1.6 3D structure of BPSL2418 _{MES}	150
5.2 Experimental structure determination of BPSL2418_{Met-SO} by molecular replacement using BPSL2418_{MES}.....	160
5.2.1 Space group determination and indexing.....	160
5.2.2 Data collection and processing.....	160
5.2.3 Matthews Coefficient (<i>V_m</i>) calculation	163
5.2.4 Structure determination of BPSL2418 _{Met-SO}	163
5.2.5 Model building and refinement.....	163
5.2.6 3D structure of BPSL2418 _{Met-SO}	167
5.3 Experimental structure determination of BPSL2418_{Reduced} by molecular replacement using BPSL2418_{MES}.....	175
5.3.1 Space group determination and indexing.....	175
5.3.2 Data collection and processing.....	175
5.3.3 Matthews Coefficient (<i>V_m</i>) calculation	178
5.3.4 Structure determination of BPSL2418 _{Reduced}	178
5.3.5 Model building and refinement.....	178
5.3.6 3D structure of BPSL2418 _{Reduced}	178
Chapter 6: BPSL2418 structure analysis.....	190
6.1 BPSL2418 structure description.....	190
6.1.1 Features of BPSL2418 structure	190
6.1.1.1 BPSL2418 _{MES}	190
6.1.1.2 BPSL2418 _{Met-SO}	200
6.1.1.3 BPSL2418 _{Reduced}	209
6.1.2 Comparison and conformational changes of MES-bound, Met-SO-bound and reduced forms of BPSL2418	214
6.1.2.1 Comparison of the overall structures of BPSL2418 forms	214
6.1.2.2 Comparison of the molecular surface of BPSL2418 forms	218

6.2 BPSL2418 interfaces, surfaces and assemblies.....	222
6.2.1 BPSL2418 _{MES} analysis	222
6.2.1.1 BPSL2418 _{MES} monomers.....	222
6.2.1.2 Monomer-monomer interface forming dimeric BPSL2418 _{MES}	222
6.2.2 BPSL2418 _{Met-SO} analysis	228
6.2.2.1 BPSL2418 _{Met-SO} monomers.....	228
6.2.2.2 Monomer-monomer interface forming dimeric BPSL2418 _{Met-SO} .	228
6.2.3 BPSL2418 _{Reduced} analysis	233
6.2.3.1 BPSL2418 _{Reduced} monomers	233
6.2.3.2 Monomer-monomer interface forming dimeric BPSL2418 _{Reduced}	233
6.3 Functional prediction.....	237
6.3.1 Structure comparisons.....	237
6.3.2 Catalytic Mechanism of BPSL2418.....	240
6.4 Conclusion and future work.....	244
Chapter 7: CT1 and CT2 Expression, purification and crystallization attempts.....	246
7.1 Preparation of CT1 protein.....	246
7.1.1 Recombinant plasmid of CT1	246
7.1.2 CT1 overexpression in <i>E. coli</i>	246
7.1.3 CT1 protein refolding.....	247
7.2 Preparation of CT2 protein.....	247
7.2.1 Recombinant plasmid of CT2	247
7.2.2 CT2 overexpression in <i>E. coli</i>	247
7.2.3 CT2 purification	250
7.2.4 Initial crystallization trials.....	254
7.2.5 Investigation of the CT2 oligomeric state by Electron Microscopy	254

7.2.6 Obtaining separated homogeneous fractions by using gel filtration....	258
7.2.7 Mutation in CT2 coiled-coil.....	258
7.2.8 Removing the His-tag	261
7.3 Conclusion and future work.....	263
Reference:.....	264

Table of Figures

Figure 1.1: Global distribution of the occurrence of melioidosis, linked to regions of <i>Burkholderia pseudomallei</i> endemicity.	27
Figure 1.2: The <i>B. pseudomallei</i> colonies.	31
(B) The organism forming a surface pellicle in selective broth.....	31
Figure 1.3: The seven morphotypes of <i>B. pseudomallei</i>	31
Figure 1.4: The structure of YKG9.....	43
Figure 1.5: The structure of PDE2A the regulatory segment of mouse.....	45
Figure 1.6: <i>E. coli</i> fRMsR structure (PDB code 1VHM).....	53
Figure 1.7: <i>Staphylococcus aureus</i> fRMsR reduction activity.....	56
Figure 1.8: Conformational changes of fRMsRred, fRMsRsub, and fRMsRox forms of <i>Staphylococcus aureus</i> fRMsR..	57
Figure 1.9: Cartoon ribbon of oxidized- binding substrate complex of <i>Neisseria meningitidis</i> fRMsR..	59
Figure 1.10: BPSL2418 is a GAF domain-containing protein.....	61
Figure 1.11: The percentages of sequence identity and similarity between the hypothetical protein from <i>Burkholderia pseudomallei</i> BPSL2418 and the GAF domain fRMsRs enzymes.	61
Figure 1.12: Amino acid Alignment of BPSL2418, <i>E. coli</i> fRMsR (1VHM), <i>Staphylococcus aureus</i> fRMsR (3KSF) and <i>Neisseria meningitidis</i> fRMsR (3MMH) proteins.	62
Figure 1.13: Cartoon representation of peptide methionine- <i>R</i> -sulfoxide reductase in <i>Burkholderia pseudomallei</i> (3CEZ) protein.	63
Figure 1.14: Kidneys affected by Polycystic Kidney Diseases.....	67
Figure 1.15: Domain structure of polycystin1.....	68
Figure 1.16: The structure prediction of the C-terminal domain of Polycystin1.....	70
Figure 1.17: Domain structure of Polycystin 2.	72
Figure 1.18: Human CT2-EF domain	73
Figure 1.19: Human CT2 coiled coil domain.....	73
Figure 1.20: The structure prediction of the C-terminal domain of Polycystin 2.....	75
Figure 2.1: pETBlue1 plasmid map.....	85
Figure 2.2: Calibration curve for gel filtration superdex 200 column.....	96
Figure 2.3: The Matrix Hydra II Plus One crystallization robot.....	99
Figure 3.1: Protein crystallization phases.....	105
Figure 3.2: The hanging drop vapour diffusion.....	107
Figure 3.3: The sitting drop vapour diffusion.....	107
Figure 3.4: This photo shows the loop that holds the crystal.	109
Figure 3.5: A Rigaku Micromax 007 copper rotating anode generator.....	112

Figure 3.6: An aerial view of Diamond Light Source, the UK's national synchrotron science facility, at the Harwell Science and Innovation Campus near Didcot, Oxfordshire..	114
Figure 3.7: Schematic drawing of Diamond Light Source.	114
Figure 3.8: Bragg's law describes the needed condition for a set of lattice planes to generate a reflection. X_0 is the incoming X-ray beam and X_1 is the diffracted beam (outgoing beam) which hits the detector.....	117
Figure 4.1: 1% DNA agarose gel of <i>bpsl2418</i> gene amplification.....	124
Figure 4.2: Blue/white screening of Transformation of ligation products into competent <i>E.coli</i> cells.	124
Figure 4.3: The <i>bpsl2418</i> insert gene was sequenced and then analyzed by BLAST, http://blast.ncbi.nlm.nih.gov/Blast.cgi	126
Figure 4.4: Confirmation of the presence and orientation of the insert gene in the pETBlue-1 plasmid.....	127
Figure 4.5: 12% SDS-PAGE gel showing the overexpression of BPSL2418.	129
Figure 4.6: A DEAE purification step showing column loading and elution of BPSL2418.	130
Figure 4.7: 12% SDS-PAGE analysis of DEAE fast flow Sepharose of BPSL2418.....	130
Figure 4.8: A Resource Q column purification step showing column loading and elution of BPSL2418.	131
Figure 4.9: 12% SDS-PAGE analysis of Resource Q column purification of BPSL2418..	131
Figure 4.10: Gel filtration purification of BPSL2418.	133
Figure 4.11: 12% SDS-PAGE analysis of purification progress of BPSL2418 preparation.	134
Figure 4.12: A 5ml Hi Trap Phenyl-HP cartridge loading and elution of BPSL2418.....	135
Figure 4.13: 12% SDS-PAGE analysis of purification progress of BPSL2418 preparation 2.....	135
Figure 4.14: photographs of initial crystallization hits of BPSL2418 _{MES} in five different conditions.	137
Figure 4.15: Photographs of BPSL2418 _{MES} crystals.....	139
Figure 4.16: Small crystals of BPSL2418 _{unbound} form produced from 0.1 M sodium acetate at pH 4.6 and 8% PEG 4000.	140
Figure 4.17: Photographs of BPSL2418 _{Met-SO} complex crystals.....	142
Figure 4.18: Photograph of crystals from BPSL2418 _{Reduced} form.....	143
Figure 5.1: X-ray diffraction image of BPSL2418 _{MES} crystal to 1.18 Å resolution using the I04 beamline of the Diamond Light Source.	147
Figure 5.2: Matthews coefficient calculations and probabilities for BPSL2418 _{MES} using Mattprob http://www.ruppweb.org/Mattprob/	149

Figure 5.3: Sample regions of electron density for BPSL2418 _{MES} show the structure improvement over 29 refinement cycles.	151
Figure 5.4: Electron density map of BPSL2418 _{MES} shows a MES molecule binding the BPSL2418 close to CYS75 and CYS109 where a disulfide bond is formed between these two cysteines.....	154
Figure 5.5: Result of MOLPROBITY and Romachandran plot of BPSL2418 _{MES} structure..	155
Figure 5.6: Main chain and side chain parameters of BPSL2418 _{MES}	156
Figure 5.7: Dimeric BPSL2418 _{MES}	158
Figure 5.8: Four BPSL2418 _{MES} molecules in a unit cell (space group P 2 ₁ 2 ₁ 2 ₁) of the crystal.....	159
Figure 5.9: X-ray diffraction image of BPSL2418 _{Met-SO} crystal to 1.4 Å using the I02 beamline of the Diamond Light Source.....	161
Figure 5.10: Matthews coefficient calculations and probabilities for BPSL2418 _{Met-SO} using Mattprob http://www.ruppweb.org/Mattprob/	164
Figure 5.11: Sample regions of electron density for BPSL2418 _{Met-SO} chain A show the structure improvement over 38 refinement cycles.....	165
Figure 5.12: Sample regions of electron density for BPSL2418 _{Met-SO} chain B show the structure improvement over 38 refinement cycles.....	166
Figure 5.13: Electron density map of BPSL2418 _{Met-SO} shows a Met-SO molecule binding the BPSL2418 close to CYS-75 and CYS-109 where a disulfide bond is formed between these two cysteines.....	169
Figure 5.14: Result of MOLPROBITY and Romachandran plot of BPSL2418 _{Met-SO} structure.....	170
Figure 5.15: Main chain and side chain parameters of BPSL2418 _{Met-SO}	171
Figure 5.16: Dimeric BPSL2418 _{Met-SO}	173
Figure 5.17: Two BPSL2418 _{Met-SO} dimers in a unit cell (space group P2 ₁) of the crystal.	
Figure 5.18: X-ray diffraction image of BPSL2418 _{Reduced} crystal to 2.0Å using the I02 beamline of the Diamond Light Source.....	176
Figure 5.19: Matthews coefficient calculations and probabilities for BPSL2418 _{Reduced} using Mattprob http://www.ruppweb.org/Mattprob/	179
Figure 5.20: Sample regions of electron density for BPSL2418 _{Reduced} chain A show the structure improvement over 18 refinement cycles.....	180
Figure 5.21: Sample regions of electron density for BPSL2418 _{Reduced} chain B show the structure improvement over 18 refinement cycles.....	181
Figure 5.22: Electron density map of BPSL2418 _{Reduced} shows no disulfide bond is formed between CYS75 and CYS109.....	183
Figure 5.23: Result of MOLPROBITY and Romachandran plot of BPSL2418 _{Reduced} structure.	185

Figure 5.24: Main chain and side chain parameters of BPSL2418 _{Reduced} . (A) Showing all the main chain parameters values are within the average for a structure at this resolution.....	186
(B) Showing all the side chain parameters values are within the average for a structure at this resolution. Produced by the programme PROCHECK (Laskowski <i>et al</i> , 1993)	187
Figure 5.25: Dimeric BPSL2418 _{Reduced} . (A) Cartoon model. (B) Space-filling model. The Figure was produced using PISA webserver http://pdbe.org/PISA	188
Figure 5.26: Two BPSL2418 _{Reduced} dimers in a unit cell (space group P2 ₁) of the crystal. (A) Spacefil model. (B) Cartoon model. The Figure was produced using PISA webserver http://pdbe.org/PISA	189
Figure 6.1: Cartoon representation of the overall fold structure of BPSL2418 _{MES}	191
Figure 6.2: Cartoon representation of the BPSL2418 _{MES} showing the active site..	192
Figure 6.3: (A) 2-(N-morpholino)ethanesulfonic acid [MES] structure with electron density map generated by COOT.....	193
Figure 6.4: The binding site interaction of BPSL2418 _{MES} with the MES buffer.....	196
Figure 6.5: Superimposition of BPSL2418 _{MES} (green) and <i>E. coli</i> fRMsr.....	197
(gray)..	197
Figure 6.6: LigPlot diagrams of the binding site interaction of <i>E.coli</i> fRMsr and BPSL2418 _{MES} . Hydrogen bonds are shown as green dashed lines between atoms..	199
Figure 6.7: Cartoon representation of the overall fold structure of BPSL2418 _{Met-SO}	201
Figure 6.8: (A) Free L-Methionine sulfoxide [free Met- <i>R</i> -SO] structure with electron density map generated by COOT. (B) The binding site of BPSL2418 _{Met-SO} with free Met-SO substrate bound.....	202
Figure 6.9: The binding site interaction of BPSL2418 _{Met-SO} with the free Met-SO substrate..	203
Figure 6.10: Superimposition of BPSL2418 _{Met-SO} (gold), <i>N. meningitides</i> fRMsr (PDB code 3MMH) (green) and <i>S. pneumonia</i> fRMsr (PDB code 3RFB) (blue)..	206
Figure 6.11: Close view of the Met-SO substrate that lies inside the active sites of BPSL2418 _{Met-SO} (gold), <i>N. meningitides</i> fRMsr (PDB code 3MMH) (green) and <i>S. pneumonia</i> fRMsr (PDB code 3RFB) (blue)..	207
Figure 6.12: LigPlot diagrams of the binding site interaction of BPSL2418 _{Met-SO} , <i>N. meningitides</i> fRMsr and <i>S. pneumonia</i> fRMsr.....	208
Figure 6.13: Cartoon representation of the overall fold structure of BPSL2418 _{Reduced}	211
Figure 6.14: Cartoon representation of BPSL2418 _{Reduced} structure showing the active site..	212

Figure 6.15: Superimposition of BPSL2418 _{Reduced} (green) chain A and <i>S. aureus</i> fRMsr (PDB code 3KSG) (light blue) chain A.....	213
Figure 6.16: Different views of the superimposition of BPSL2418 _{MES} (gold), BPSL2418 _{Met-SO} chain A (blue) and BPSL2418 _{Reduced} chain A (green) with disulfide bonds in yellow..	215
Figure 6.17: Superimposition of the three forms of BPSL2418 shows the change in the active site, BPSL2418 _{MES} (gold), BPSL2418 _{Met-SO} (blue) and BPSL2418 _{Reduced} (green)..	216
Figure 6.18: Calpha-Calpha distance (Y-axis) of residues of loop4 between BPSL2418 _{Met-SO} and BPSL2418 _{Reduced} (blue line) and between BPSL2418 _{MES} and BPSL2418 _{Reduced} plotted against residues number of loop4 (X-axis).	217
Figure 6.19: B-factors (Å ²) of main chain residues of BPSL2418 different forms plotted against residues numbers (X-axis).....	219
Figure 6.20: Comparison of electrostatic surface potential of BPSL2418 forms showing the conformational changes in the active site.....	221
Figure 6.21: The dimeric structure model of BPSL2418 _{MES}	223
Figure 6.22: Residues positioned in BPSL2418 _{MES} monomer-monomer interface.....	227
Figure 6.23: The dimeric structure model of BPSL2418 _{Met-SO}	229
Figure 6.24: Residues positioned in BPSL2418 _{Met-SO} monomer-monomer interface....	232
Figure 6.25: Residues positioned in BPSL2418 _{Reduced} monomer-monomer interface...236	
6.26: Structure based alignment of BPSL2418 (s001), <i>N. meningitidis</i> fRMsr (s004), <i>S. pneumoniae</i> fRMsr (s005), <i>E. coli</i> fRMsr (s006) and <i>S. aureus</i> fRMsr (s007). The β-strands and α-helices are represented as blue and red residues, respectively....	238
6.27: structural alignment of BPSL2418 structure and fRMsr structures (<i>N. meningitidis</i> fRMsr <i>S. pneumoniae</i> fRMsr, <i>E. coli</i> fRMsr and <i>S. aureus</i> fRMsr).....	239
Figure 6.28: Structural representation of proposed catalytic mechanism of BPSL2418..	242
Figure 6.29: Schematic representation of proposed catalytic mechanism of BPSL2418.	243
Figure 7.1: 12% SDS- PAGE gel showing GB1-CT1 (34 kDa) expression.....	248
Figure 7.2: His-CT2 construct (35 kDa).....	249
Figure 7.3: 12% SDS- PAGE gel showing His-CT2 (35 KDa) overexpression.....	251
Figure 7.4: 12% SDS-PAGE of His-CT2 (35 KDa) nickel affinity chromatography.....	252
Figure 7.5: Gel filtration of His-CT2 (35 KDa).....	253
Figure 7.6: The calibration plot for Hi-Load superdex 200 column chart to calculate CT2 molecular weight.....	255
Figure 7.7: Quasi crystals of CT2 grew in 0.2 M ammonium phosphate, 0.1 M Tris pH 8.5 and 50% MPD.....	256
Figure 7.8: EM image for the oligomeric state of His-CT2.....	257

Figure 7.9: Using gel filtration to achieve homogenous fractions of His-CT2.	259
Figure 7.10: Mutation in CT2 coiled-coil.....	260
Figure 7.11: Removing the His-tag experiment.....	262

List of Tables

Table 5.1: Data collection statistics for the BPSL2418_{MES}.....	148
Table 5.2: Refinement statistics for BPSL2418_{MES} after 29 cycles of rebuilding in COOT and refinement in REFMAC5.....	152
Table 5.3: Data collection statistics for the BPSL2418_{Met-SO}.....	162
Table 5.4: Refinement statistics for BPSL2418_{Met-SO} after 38 cycles of rebuilding in COOT and refinement in REFMAC5.....	168
Table 5.5: Data collection statistics for the BPSL2418_{Reduced}	177
Table 5.6: Refinement statistics for BPSL2418_{Reduced} after 18 cycles of rebuilding in COOT and refinement in REFMAC5.....	184
Table 6.1: Dali server result for the model BPSL2418_{MES}.....	195
Table 6.2: Dali server result for the model BPSL2418_{Met-SO}.....	205
Table 6.3: Dali server result for the model BPSL2418_{Reduced}.....	210
Table 6.4: Accessibility and buried surface areas of 27 residues involved in the BPSL2418_{MES} monomer-monomer interface calculated by PISA.....	225
Table 6.5: The hydrogen bonds and salt bridges formed in BPSL2418_{MES} monomer-monomer interface calculated by PISA.....	226
Table 6.6: Accessibility and buried surface areas of 25 residues involved in the BPSL2418_{Met-SO} monomer-monomer interface calculated by PISA.....	230
Table 6.7: The hydrogen bonds and salt bridges formed in BPSL2418_{Met-SO} monomer-monomer interface calculated by PISA.....	231
Table 6.8: Accessibility and buried surface areas residues involved in the BPSL2418_{Reduced} monomer-monomer interface calculated by PISA.....	234
Table 6.9: The hydrogen bonds and salt bridges formed in BPSL2418_{Reduced} monomer-monomer interface calculated by PISA.....	235

Abbreviations

Biological and chemical terms	
A	Adenine
ADPKD	Autosomal Dominant Polycystic Kidney Disease
AHLs	<i>N</i> -Acyl homoserine lactones
ALA	Alanine
Amp	Ampicillin
ARG	Arginine
ARPKD	Autosomal Recessive Polycystic Kidney Disease
ASN	Asparagine
ASP	Aspartic acid
BPSL	<i>Burkholderia pseudomallei</i> Large chromosome
BPSS	<i>Burkholderia pseudomallei</i> Small chromosome
BPSL2418_{MES}	The BPSL2418 MES-bound form
BPSL2418_{Met-SO}	The BPSL2418 Met-SO-bound form
BPSL2418_{Reduced}	The BPSL2418 Reduced form
C	Cytosine
Cam	Chloramphenicol
cAMP	Cyclic adenosine monophosphate
Car	Carbenicillin
cDNA	complementary DNA
CFE	Cell free extract
cGMP	Cyclic guanosine monophosphate
cNMP	Cyclic nucleotide monophosphate
CYS	Cysteine
CT1	The C-terminal of Polycystin 1
CT2	The C-terminal of Polycystin 2
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EPS	Extracellular polymeric substance
fRMsr	Free methionine- <i>R</i> -sulfoxide reductase
G	Guanine
GAF	cGMP-specific phosphodiesterases, adenylyl cyclases and FhlA
GB1	Immunoglobulin-binding fusion
GLN	Glutamine

GLU	Glutamic acid
GLY	Glycine
HeLa cell	Helen Larson cell
HIS	Histidine
Kan	Kanamycin
Ig-like PKD	Immunoglobulin like polycystic kidney disease
LD₅₀	Lethal dose 50%
LDL-A	Low-density lipoprotein A
LEU	Leucine
LPS	Lipopolysaccharides
LYS	Lysine
MES	2-(<i>N</i> -morpholino)ethanesulfonic acid
MET	Methionine
Met-SO	Methionine sulfoxide
Met-SO₂	Methionine sulfone
Met-R-O	The R form of methionine sulfoxide
Met-S-O	The S form of methionine sulfoxide
Msr	Methionine sulfoxide reductase
NADPH	Nicotinamide Adenine Dinucleotide Phosphate Hydrogen
NHS	Normal human serum
PAGE	Polyacrylamide gel electrophoresis
PC1	Polycystin1
PC2	Polycystin2
PDEs	Phosphodiesterases
PEG	Polyethylene glycol
PHE	Phenylalanine
PKA	Protein kinase A
PKD1	Polycystic kidney disease1
PKD2	Polycystic kidney disease2
PKG	Protein kinase G
PMSF	phenylmethylsulfonyl fluoride
PRO	Proline
REJ	Receptor for egg jelly
ROS	Reactive oxygen species
SDS	Sodium dodecyl sulfate
SER	Serine
T	Thymine
TAE buffer	A mixture of Tris base, acetic acid and EDTA

TEMED	Tetramethylethylenediamine
Tet	Tetracycline
THR	Threonine
Tris	tris(hydroxymethyl)aminomethane)
TRP	Tryptophan
TRP	Transient receptor potential channel
Trx	Thioredoxin
T2SS	Type II general secretory system
T3SS	Type III general secretory system
T6SS	Type VI general secretory system
TYR	Tyrosine
VAL	Valine
X-gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside
Crystallographic terms	
Å	Angstrom (10^{-10} m)
ASU	Asymmetric unit
a, b, c, α, β, γ	Real space unit cell dimensions and angles
B factor	Crystallographic temperature factor
d_{hkl}	The distance between planes with indices $h k l$ in the reciprocal lattice
hkl	Reciprocal lattice indices
F	Structure factor
F_{hkl}	Structure factor for a single reflection with indices hkl
F_{hkl}	Structure factor amplitude for the reflection with indices hkl
F_{calcs}	Calculated structure factor
F_{obs}	Observed structure factor
F_{calcs}	Calculated structure factor amplitude
F_{obs}	Observed structure factor amplitude
f	Atomic scattering factor
$I / \sigma I$	Signal to noise ratio
MAD	Multi wavelength anomalous dispersion
MR	Molecular replacement
R-factor	Crystallographic refinement R-factor
R_{free}	Free R-factor
R_{merge}	Merging R-factor
R_{pim}	Precision-indicating merging R-factor
SAD	Single-wavelength anomalous dispersion
u, v, w	co-ordinates in Patterson space
V_m	Matthews Coefficient

XDS	X-ray Detector Software
Z	Number of equivalent positions in the unit cell
ρ	Electron density
λ	X-ray wavelength
Miscellaneous terms	
°	Degrees
%	percentage
BLAST	Basic local alignment search tool
°C	Degrees centigrade
CCP4	Collaborative Computational Project No. 4
3D	Three dimensional
EM	Electron microscopy
h	hour
kDa	Kilodalton
LB	Luria Bertani media
MW	Molecular weight
NMR	Nuclear magnetic resonance
OD	Optical density
PCR	polymerase chain reaction
PI	Isoelectric point
RMSD	Root-mean-square deviation
rpm	Revolutions per minute
SOC	Super optimal broth with catabolite repression
UV	Ultraviolet

Chapter 1: Introduction

This thesis consists of two projects. The first and most important part in terms of achieved results is BPSL2418 the hypothetical protein from *Burkholderia pseudomallei*. The second project is the C-terminal domains of Polycystin 1 (CT1) and Polycystin 2 (CT2).

First project: BPSL2418 from *Burkholderia pseudomallei*

This chapter gives a general introduction of the hypothetical protein from *Burkholderia pseudomallei* (BPSL2418). It starts with a general view of the bacterial pathogen *Burkholderia pseudomallei* and its biological relevance as the causative agent of the human disease melioidosis. Then it focuses on the family of proteins to which BPSL2418 belongs. Finally it states the aims of this project.

1.1 *Burkholderia pseudomallei* is the causative agent of melioidosis

1.1.1 Background and history

Melioidosis, is an infection caused by the Gram-negative bacterium *Burkholderia pseudomallei*, first identified by the British pathologist Captain Alfred Whitmore and his assistant C S Krishnaswami in Rangoon, Burma in 1911 (White, 2003b). They described melioidosis as a glanders-like disease among the ill residents of Rangoon. Glanders is an abscess-forming infection caused by *Burkholderia mallei* which affects horses, mules and donkeys although it can be contracted by other animals such as dogs, cats and goats and it occasionally affects humans (Bondi & Goldberg, 2008). Whitmore and Krishnaswami isolated the causative agent of

melioidosis from autopsy specimens on peptone agar and potato slopes, they identified a new organism that corresponding to Koch's postulates (Cheng & Currie, 2005b). This organism is distinguished from the organism causing glanders by its comparatively rapid growth, its motility and the lack of the Strauss reaction when it was injected into guinea pigs (Cheng & Currie, 2005b). In 1932 Stanton and Fletcher (from the Institute of Medical Research in Kuala Lumpur, Malaya) documented sufficient cases in man and animals of Whitmore's disease to publish a definitive monograph on a disease they named as melioidosis (White, 2003b). The name melioidosis was derived from two Greek words [*melis*]: which means a distemper of asses and [*eidōs*]: resemblance (Dance, 1991a). After the Second World War many more cases were reported in American and western soldiers fighting in Vietnam, who had been exposed to environmental *Burkholderia pseudomallei* through contaminated wounds and burns or by inhalation (Dance, 1991a). The organism causing melioidosis has been known as *Bacillus pseudomallei*, *Bacillus whitmorii* (or *Bacille de whitmore*), *Malleomyces pseudomallei* and *Pseudomonas pseudomallei* but since 1992 it has been termed *Burkholderia pseudomallei* (Cheng & Currie, 2005b). However, melioidosis appeared as an infectious disease of major public health importance in Southeast Asia and northern Australia in the latter half of the 20th century (Chaowagul *et al*, 1989).

1.1.2 Epidemiology

Melioidosis is considered as endemic to the tropical area of Southeast Asia and Northern Australia; multiple case series have been reported and described in Northern Australia, Thailand, Singapore, Malaysia, Burma and Vietnam (Dance, 2000b) (Cheng & Currie, 2005b). Significant numbers of exported cases have occurred in countries classified as possibly endemic regions such as Southern China, Hong Kong, Brunei, Laos, Cambodia and Taiwan (Figure 1.1) (Cheng & Currie, 2005). Also a limited outbreak of the disease was reported in Europe. The most remarkable occurrence of melioidosis was in France in the mid-1970s: this outbreak started in Paris zoo when a panda infected with the bacteria was donated

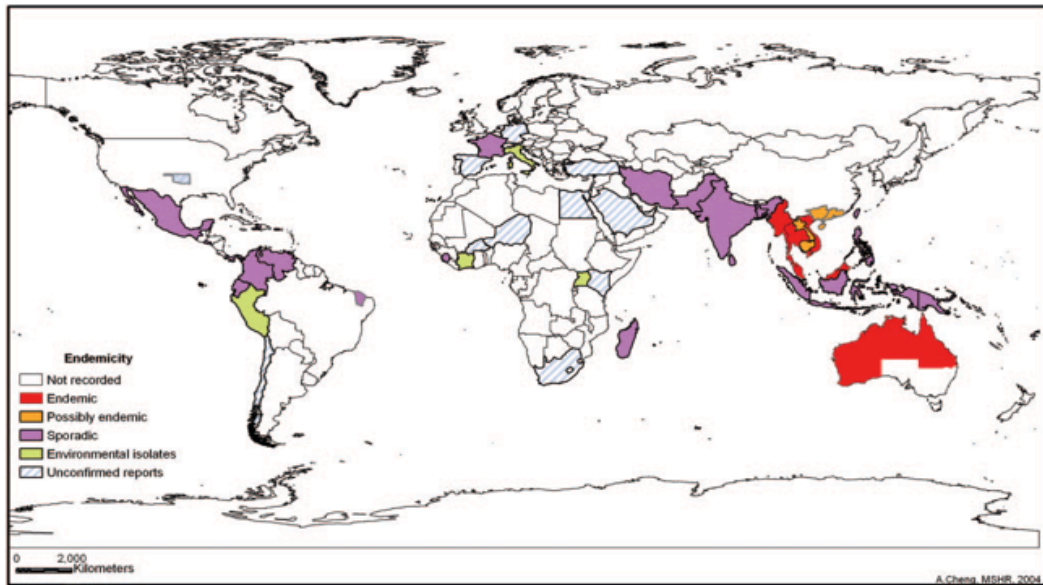


Figure 1.1: Global distribution of the occurrence of melioidosis, linked to regions of *Burkholderia pseudomallei* endemicity. Adapted from (Cheng & Currie, 2005b)

from China, and subsequently the infection distributed to other zoos in and outside France, probably by transport of affected animals or contaminated muck. An unknown number of slaughtered animals and at least two fatal human cases resulted from this infection (Dance, 1991a). In Spain an isolation of *Burkholderia pseudomallei* from horses was reported. Also some human melioidosis cases have been announced in Berlin in 1947 (Dance, 1991a).

Sporadic cases have appeared in Central and South America, the Middle East, the Pacific and several African countries, and at least five cases in the United Kingdom imported to the country from Bangladesh and two from India and Pakistan (Cheng & Currie, 2005). However, the worldwide epidemiology of melioidosis shows that, Northeast Thailand has the highest incidence of melioidosis where the first case wasn't reported until 1955, but since 1986 the MORU lab (Mahidol-Oxford Tropical Medicine Research Unit) in Bangkok has documented more than 3000 cases (Stone, 2007) and Sappasitprasong Hospital in Ubon Ratchathani in north-east Thailand admits around 200 patients annually that are positive for *Burkholderia pseudomallei* infection, of which nearly half die (Aldhous, 2005). *Burkholderia pseudomallei* is found in the environment of the endemic countries where the highest isolation rate was in wet rice fields in Malaysia and Thailand, but in Australia the highest rate was in dry soil. However, this bacterium exists in soil and water in all endemic countries (Dance, 2000a; Dance, 2000b). Although several studies have spotted melioidosis, it is thought that the extent of the disease in the endemic area as well as worldwide is misrepresented, with many unreported cases due to lack of awareness and diagnostic facilities in infected regions, especially in rural communities. Melioidosis is categorized as an emerging infection, because it has been recognized increasingly during the past two decades (Aldhous, 2005; Dance, 2000b).

1.1.3 *Burkholderia pseudomallei*

1.1.3.1 Bacterial strain and genomics

The genus *Burkholderia* includes more than 30 species where the most pathogenic

types are *B. pseudomallei*, *B. mallei* and *B. thailandensis*, which exists with *Burkholderia pseudomallei* in the soil in Thailand but is less virulent (Cheng & Currie, 2005b). The *Burkholderia pseudomallei* strain K96243 was isolated in 1996 from a 34-year-old female diabetic patient in Khon Kaen hospital in Thailand, and then the DNA was isolated from this strain (Holden *et al*, 2004). The *B. pseudomallei* genome possesses two chromosomes, of which BPSL (*Burkholderia pseudomallei* Large) is 4.07 megabase pairs and BPSS (*Burkholderia pseudomallei* Small) is 3.17 megabase pairs. The large chromosome regulates metabolism and growth, whereas the small chromosome has more accessory functions relating to adaptation and survival in various circumstances (Holden *et al*, 2004). A further astonishing aspect of the *B. pseudomallei* genomic structure is the presence of 16 genomic islands that compose 6.1% of the genome, but are completely missing in the paraphyletic non-human pathogenic *B. mallei* and *B. thailandensis*, or in the plant pathogenic *B. cenocepacia*. This suggests a significant role for this genomic island in the pathogenicity of *B. pseudomallei* to humans and animals ADDIN EN.CITE .

1.1.3.2 Bacteriology

Burkholderia pseudomallei is characterized as a Gram-negative aerobic bacillus with bipolar staining and has a slender, rounded end shape. It has been described as safety-pin like (Cheng & Currie, 2005b; White, 2003a). *B. pseudomallei* is a soil saprophyte, lives in water and wet soils in rice paddy fields in endemic regions, and it can be cultured from more than 50% of rice paddies (Cheng & Currie, 2005b; Dance, 2000a; White, 2003a). The bacterium is motile where Whitmore distinguished it from *B. mallei* by its ability to move on a hanging drop (Cheng & Currie, 2005b). The optimum growth condition of *B. pseudomallei* is in paddy soil, in crop-covered and dormant fields. It flourishes in neutral to acidic medium (pH 6.5 or 7.5) but not in salty water or soil (Chen *et al*, 2003). 37 °C or 42 °C are the optimum temperatures for growth although this bacterium can still grow at 4 °C which explains the existence of melioidosis in some cold areas. Also the water content of the medium is a critical element influencing the degree of growth (Chen *et al*, 2003). However the organism is highly resistant to difficult environmental

conditions and it is able to survive in a wide range of harsh circumstances in both soil or liquid media (Cheng & Currie, 2005b). The bacteria remain active and continue to grow under a diverse range of temperature, pH and dehydration conditions (Chen *et al*, 2003), they also conserve vitality after at least 16 years in distilled water (Pumpuang *et al*, 2011). Moreover they can survive even if exposed to disinfectant and antiseptic solutions (Gal *et al*, 2004) and they show a resistance to chlorinated water with levels used in water supplies proving only bacteriostatic, although use of high levels are sufficient for killing the bacteria (Howard & Inglis, 2003). The organism is oxidase positive, where the enzyme is able to assimilate arabinose or glucose in the oxidative pathway, this also allows the distinction of *B. pseudomallei* from the closely related but less pathogenic *B. thailandensis* (Cheng & Currie, 2005b; White, 2003a). In the laboratory, *B. pseudomallei* can be cultured on most agar media, and colonies can be obviously seen within 24 h at 37 °C, although usually Ashdown's selective medium is used for growing the organism where the *B. pseudomallei* colonies have a cornflower head appearance and they are stained with the crystal-violet dye from the Ashdown's media (White, 2003b) (Figure 1.2). Isolates of the bacteria from environmental sources or from infected patients show various morphologies with seven colony types (Figure 1.3) where Type I is the most commonly isolated morphotype of the seven types (Chantratita *et al*, 2007). A wide range of morphologies further obscures identification of melioidosis upon culturing from infected patients (Stone, 2007).

1.1.4 Melioidosis features

1.1.4.1 Melioidosis acquisition

B. pseudomallei is a soil saprophyte and can be cultured from soil and surface water in endemic regions, with postulated routes of infection being via contamination of broken skin and wounds, aspiration, inhalation of aerosolized bacteria, ingestion and possibly sexual transmission (Currie *et al*, 2001). Mother-to-child transmission of *B. pseudomallei* may also possibly have occurred as a result of placental infection (Abbink *et al*, 2001).

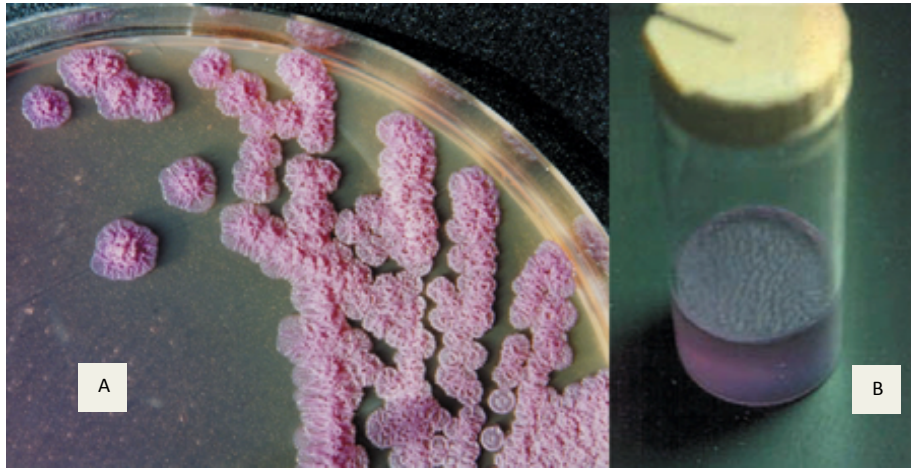


Figure 1.2: The *B. pseudomallei* colonies. (A) *B. pseudomallei* growing on Ashdown's selective medium.

(B) The organism forming a surface pellicle in selective broth. Adapted from (White, 2003b)

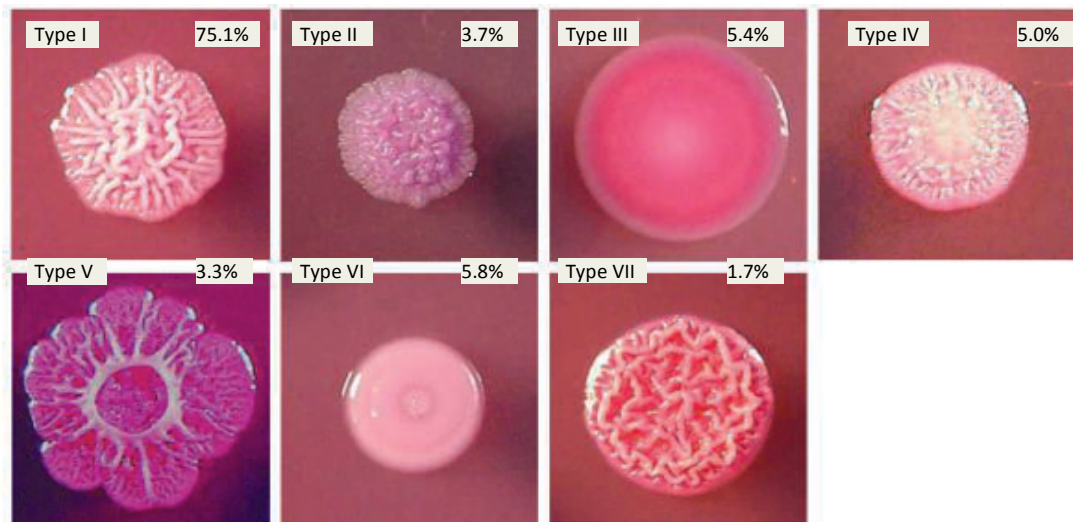


Figure 1.3: The seven morphotypes of *B. pseudomallei*. Colonies were morphotyped based on appearance, size and pigmentation. The percentages indicate the proportion of each form in a collection of 212 clinical isolates after growth for four days on Ashdown's agar. Adapted from (Chantratita *et al*, 2007; Stone, 2007).

1.1.4.2 Melioidosis symptoms

Melioidosis is known as “the great mimicker” because of its similar symptoms to other infections and the difficulty with its diagnosis (Yee *et al*, 1988). This disease can be classified into two categories, an acute septicaemia infection and a chronic localized infection (Dance, 1991a). It can affect any organ or tissue but the common positions of infection include lungs, genito-urinary system, skin, liver, spleen, bones, soft tissues and skeletal muscle (Yee *et al*, 1988). Symptoms may include pain in chest, bones, joints, coughing, skin infections, lung nodules and pneumonia, the most common manifestation of melioidosis (Suputtamongkol *et al*, 1999). Melioidosis occurs as an acute bloodstream infection, disseminated infection and most frequently as an acute pulmonary infection. Chronic melioidosis may present as a localized infection in the skin, liver, spleen, kidneys or prostate, but also it occurs as chronic pneumonia, visceral abscesses and osteomyelitis (Lim *et al*, 2001). The incubation period is generally 1-21 days, but may extend to months or years, symptoms appear 2-4 weeks after exposure. Without appropriate treatment, the case-fatality ratio may reach 90% within 48 hours of developing symptoms (Cheng & Currie, 2005a).

1.1.4.3 Environmental exposure

Melioidosis affects humans and animals that have been exposed to environments including *B. pseudomallei*. The mode of infection is considered to be either through cutaneous inoculation, inhalation, ingestion or aspiration (Pumpuang *et al*, 2011). The connection between the disease prevalence and the degree of environmental contamination with *B. pseudomallei* is a direct relation where the highest disease incidence is in the environments of the endemic regions that contain a very high level of the bacteria (Cheng & Currie, 2005b).

1.1.4.4 Climate conditions

The weather is another factor associated with the disease acquisition, where heavy rain, monsoon and winds trigger the highest number of cases (Dance, 2000a). This can be illustrated by rising opportunities for contamination with *B. pseudomallei*

where the bacteria is widely spread in pooled surface water such as in rice paddies (Cheng & Currie, 2005b). The rice farmers in endemic areas plant at the start of the monsoon and work in flooded rice paddies until harvest in very low conditions of safety, they are always barefoot in rice fields, which explains the high risk of infection distribution in this kind of agricultural activity (Stone, 2007).

1.1.4.5 The host immunity status

One fifth of all melioidosis patients in northeast Thailand (80% of the population belong to rice-farming families) are children under 14 years of age, while the disease extent peak includes individuals between the fourth and sixth decade (Wiersinga *et al*, 2007). Most of them have other diseases, commonly diabetes mellitus, renal failure and thalassemia (Cheng & Currie, 2005b). There is a possibility that the development of these diseases causes immune dysfunction, which increases the chance of melioidosis infection to take place (Cheng & Currie, 2005b; Wiersinga *et al*, 2007). Also it is uncertain whether the afflicted children with melioidosis have a major genetic tendency for disease. It also likely that melioidosis in childhood occurs due to the high bacterial distribution which increases the pathogenic capability (Wiersinga *et al*, 2007).

1.1.4.6 Treatment of melioidosis

The current therapy for melioidosis is similar to that of any Gram-negative septicaemia. The treatment consists of two stages, primary high intensity intravenous injections and oral treatment (White, 2003b). The primary treatment includes a high dose of ceftazidime for at least ten days or until progress appears, followed by oral treatment of a cocktail of four conventional antibiotics including chloramphenicol, doxycycline, trimethoprim and co-trimoxazole for 20 weeks. With the four-drug combination, chloramphenicol is given only for the first 8 weeks (White *et al*, 1989). Without treatment the fatality rate of acute melioidosis is over 80%, the primary intravenous stage itself decreases overall mortality more than 50% (White *et al*, 1989). Reversion into melioidosis occurs in approximately 10% of patients due to the reactivation of the original infection (White, 2003b).

1.1.4.7 New therapies and vaccines are needed

Burkholderia infections are inherently difficult to treat due to their resistance to multiple antibiotics, biofilm formation, and the establishment of intracellular and chronic infection in the host (Wiersinga *et al*, 2006). Until recently, the four-drug combination of antibiotics has been the only available treatment for melioidosis but it is an imperfect remedy due to the extended time and cost of an antibiotics course with the fatality rate remaining high (Aldhous, 2005; Behera *et al*, 2012), and there is also the rise of a number of ceftazidime resistant strains (Behera *et al*, 2012). Preventive factors such as active immunization might significantly decrease the risk of disease incidence, but to date there is no commercially available effective vaccine that prevents *B. pseudomallei* infection, although several approaches have been taken to develop effective vaccines for use against melioidosis (Peacock *et al*, 2012). Different live attenuated mutants have been used as vaccines in mouse models providing high levels of infection resistance (Nieves *et al*, 2011). The use of *B. pseudomallei* as a potential vaccine has been tested showing successful protection in mice but was found to be ineffective in human trials (AuCoin *et al*, 2012). Also the use of monoclonal antibodies specific to the bacteria lipopolysaccharides and capsular polysaccharide for passive immunization has been shown to offer protection from infection in mice (AuCoin *et al*, 2012). But unfortunately these vaccines are unlikely to be developed for human use due to the high chance of reversion and the ability of the *B. pseudomallei* to start a persistent latent infection, which can remain dormant for extended periods of time (Nieves *et al*, 2011).

1.1.4.8 Melioidosis as a potential bioweapon

The United States Center for Disease Control considers *B. pseudomallei* as a class B potential biological threat which categorizes the organism in the second highest priority level among other potential bioweapons (Rotz *et al*, 2002). The disease is placed in this critical position due to the high mortality rate, the simplicity of disease acquisition and spreading, the ability of the bacteria to survive in a wide range of conditions, the lack of vaccine and the multi antibiotic resistance of *B. pseudomallei* (Cheng & Currie, 2005a).

1.1.5 Molecular pathogenicity factors

1.1.5.1 Quorum sensing

Quorum sensing is a cell-density-dependent communication system observed in Gram-negative bacteria that permits bacteria to regulate gene expression using the small signaling molecules *N*-acyl-homoserine lactones (AHLs) (Gamage *et al*, 2011). The *B. pseudomallei* genome contains three LuxI and five LuxR genes responsible for producing eight quorum sensing homologous LuxI and LuxR proteins. The LuxI proteins play role in AHLs biosynthesis, and the LuxR proteins act as transcriptional regulators (Gamage *et al*, 2011). This system regulates the expression of a number of genes involving metalloproteases and phospholipase C, and also it is responsible for biofilm formation (Gamage *et al*, 2011; Wiersinga *et al*, 2007). Interruption of the LuxI or LuxR genes requires a significant increase in the LD₅₀, increases the lifetime and decreases bacterial colonization and virulence in Syrian hamsters, which suggests that the synthesis of *N*-acyl-homoserine lactones (AHLs) is critical for full virulence in this rodent model (Ulrich *et al*, 2004). Also there is proof for the requirement of the Bpe-OprB multidrug efflux pump for AHLs to be released from inside the cell (Chan & Chua, 2005a). The Bpe-OprB multidrug efflux pump in *B. pseudomallei* also is implicated in the antimicrobial resistance to aminoglycosides and macrolides (Wiersinga *et al*, 2007).

1.1.5.2 Capsular polysaccharides

Pathogenic bacteria commonly produce a thick and mucous-like layer of polysaccharide that forms a capsule outside the cell wall. The capsule covers antigenic proteins on the bacterial surface that stimulate an immune response. It protects the cell from phagocytosis where a capsule-specific antibody may be required for phagocytosis to take place. However it is known that bacterial capsules are available in several different types and immunity to one type does not result in immunity to the other types. Also this layer assists the cell to adhere to surfaces. The bacterial capsule regularly consists of polysaccharides although it can be

composed of polypeptides (O'Riordan & Lee, 2004). *Burkholderia pseudomallei* is morphologically distinguished into three types based on its polysaccharide capsule (Puthuchearry *et al*, 1996). One type has no capsule while the other two types have various capsule thicknesses. The *B. pseudomallei* possesses a capsule with a homopolymer of structure -3)-2-O-acetyl-6-deoxy- β -D-manno-heptopyranose-(1- (Reckseidler *et al*, 2001). The presence of polysaccharide capsule is significant for *B. pseudomallei* virulence in experimental animal models. The addition of purified *B. pseudomallei* capsule to a serum bactericidal test boosts the survival of *B. pseudomallei* SLR5 (Reckseidler *et al*, 2001). Also an increase in the numbers of wild-type *B. pseudomallei* cells in the blood was shown after 48 h while the number of capsule-deficient mutant cells in the blood decreased after 48 h (Reckseidler-Zenteno *et al*, 2005).

1.1.5.3 Lipopolysaccharide

Lipopolysaccharides (LPS) are macromolecules composed of lipid and polysaccharide bound by a covalent bond. It is a major component of the outer membrane of Gram-negative bacteria, functions as an endotoxin and plays a role in immune responses in animals (Nikaido & Vaara, 1985). *Burkholderia pseudomallei* owns an individual LPS structure with an unbranched repeating unit of alternating glucose and talose residues in the form -3)- β -D-glucopyranose-(1-3)-6-deoxy- α -L-talopyranose-(1- with the talose subunits acetylated on the 2' and 3' positions (Perry *et al*, 1995). *B. pseudomallei* is highly resistant to the bactericidal activity of normal human serum (NHS) where the bacteria multiplies in 10-30% NHS. But it was found that *B. pseudomallei* mutants in the LPS synthetic pathway are debilitated in rodent models with the bacteria becoming susceptible to be killed by the alternative complement pathway and to be less virulent (Deshazer *et al*, 1998).

1.1.5.4 Flagella

B. pseudomallei has a polar tuft of two to four flagellae that supply motility but the flagella's role in virulence is still obscure due to conflicting data. Flagella-deficient

B. pseudomallei mutants are show a reduction in numbers in lung and spleen of mice as compared with the wild type (Wiersinga *et al*, 2007). But in another study there was no difference between the wild type and the mutant when invading and replicating in human lung cell *in vitro* (Chua *et al*, 2003).

1.1.5.5 Pili

Adherence is a critical virulence factor in many Gram-negative bacteria, mediated by carbohydrate molecules, pilus and non-pilus adhesion. The *Burkholderia pseudomallei* K96243 genome includes multiple type IV pilin-associated loci, including the *pilA* gene, which is responsible for producing a putative pilus structural protein (Essex-Lopresti *et al*, 2005). Mutations (deletion) in the *pilA* gene reduce the bacterial adherence to human epithelial cells *in vitro* and they also decrease bacterial virulence in nematode models, indicating a role for type IV pili in *Burkholderia pseudomallei* virulence (Essex-Lopresti *et al*, 2005; Wiersinga *et al*, 2007).

1.1.5.6 Biofilm formation

A biofilm is a group of microorganism in which cells adhere to each other on a surface. These sticking cells are encased in an extracellular polymeric substance (EPS) of microbial origin. Biofilm (EPS) is described as slime consisting of extracellular DNA, proteins and polysaccharides (Costerton *et al*, 1999). *Burkholderia pseudomallei* is capable of forming biofilms (Vorachit *et al*, 1995). Biofilm formation increases the cells' antibiotic resistance (Sawasdidoln *et al*, 2010), but mutants lacking biofilms are not attenuated in mouse models, suggesting it is not essential for virulence (Taweechaisupapong *et al*, 2005). However the embedding of bacteria in biofilms in infected lung tissue indicates a function in bacteria survival, vitality and antibiotic resistance inside the host (Sawasdidoln *et al*, 2010).

1.1.5.7 Antibiotic resistance

Burkholderia pseudomallei is resistant to a wide range of antibiotics such as β -lactams, aminoglycosides, macrolides and polymyxins (White, 2003b). The genome contains genes encoding seven β -lactamases providing resistance against a series of β -lactam based antibiotics which includes penicillin derivatives (penams), cephalosporins (cephems), monobactams, and carbapenems. Most β -lactam antibiotics work by inhibiting cell wall biosynthesis in the bacterial organism. Bacteria usually develop resistance to β -lactam antibiotics by producing β -lactamases (White, 2003b). Also the *B. pseudomallei* genome encodes a putative aminoglycoside acetyltransferase, which is responsible for some of the bacteria's resistance to aminoglycosides alongside other systems (White, 2003b).

In addition to the antibiotic resistance enzymes, *B. pseudomallei* produces a number of multi-drug efflux pumps, seven putative systems of unknown function and three that are fully characterized, AmrAB-OprA (BPSL1802 – BPSL1805), BpeAB-OprB (BPSL0813 – BPSL0816) and BpeEF-OprC (BPSS0290 – BPSS0294) (Kumar *et al*, 2008). The AmrAB-OprA efflux pump is a distinguishing system for the extraction of aminoglycosides and macrolides (Moore *et al*, 1999). The BpeAB-OprB multi-drug efflux pump also provides antimicrobial resistance by the extrusion of aminoglycosides and macrolides (Chan & Chua, 2005b). Most *Burkholderia pseudomallei* strains are intrinsically aminoglycoside resistant, mainly due to AmrABOprA and BpeAB-OprB multi-drug efflux pumps, but in a single separate study the data show that BpeAB-OprB from 1026b strain does not mediate efflux of aminoglycosides, but is a multidrug efflux system which extrudes macrolides, fluoroquinolones, tetracyclines, acriflavine, and, to a lesser extent, chloramphenicol (Mima & Schweizer, 2010). The BpeEF-OprC efflux pump is responsible for the extrusion of chloramphenicol and trimethoprim (Kumar *et al*, 2008).

1.1.5.8 Secretion systems

Clinical isolates of *Burkholderia pseudomallei* from humans were tested for their ability to produce extracellular, biologically active substances, which are thought to contribute to the virulence of bacteria. *B. pseudomallei* is able to produce lipases, proteases, haemolysins and siderophores (Ashdown & Koehler, 1990). Different

strains of *B. pseudomallei* show various levels of proteolytic activity in cell free supernatants, but nevertheless there is no correlation between protease activity and virulence in mouse models of infection (Gauthier *et al*, 2000). The protease enzymes are capable of huge damage to cell lines, localized tissue harm and necrosis if injected into guinea pig or rabbit models (Ling *et al*, 2001). Also *B. pseudomallei* produces phospholipase C enzymes, which are found to contribute to cytotoxicity of HeLa cells (Korbsrisate *et al*, 2007). It is also able to produce siderophores high-affinity iron chelating compounds, which increase iron uptake into the cell (Yang *et al*, 1991).

1.1.5.8.1 Type II secretion system

It has been demonstrated in several studies that *B. pseudomallei* secretes protease, lipase and phospholipase C into the extracellular environment. The bacteria possess a type II general secretory system (T2SS) that is able to secrete the above-mentioned proteins (DeShazer *et al*, 1999). However T2SS mutants are not significantly debilitated in hamster models, indicating a minor role for these proteins in pathogenicity (Gauthier *et al*, 2000).

1.1.5.8.2 Type III secretion system

The *B. pseudomallei* contains three type III secretion system T3SS (T3SS1, T3SS2 and T3SS3). T3SS1 and T3SS2 are homologous to the T3SS of the plant pathogen *Ralstonia solanacearum*, while T3SS3 shows homology to the T3SS of pathogenic *Salmonella* species. The T3SS secretion apparatus functions like a molecular syringe. A subset of type III proteins (translocators) interact with the eukaryotic cell membrane and inject other type III proteins (effectors) into the target-cell cytosol, where they destroy host-cell processes (Wiersinga *et al*, 2006). T3SS3 has been shown to have a role in survival and persistence inside macrophages, escape from endocytic vesicles, multinucleated giant cell formation and induction of host cell apoptosis (Stevens *et al*, 2002) (Suparak *et al*, 2005). An initial study using deletion mutants of the SctU subunit, a major component of the inner membrane assembly for the three systems found only T3SS3 to be an important pathogenicity determinant in mouse and hamster models (Warawa & Woods, 2005).

1.1.5.8.3 Type VI secretion system

One of the factors that offer *B. pseudomallei* the ability to survive and multiply within a variety of eukaryotic cells, including macrophages, is the presence of six type VI secretion systems (T6SS I, T6SS II, T6SS III, T6SS IV, T6SS V, T6SS VI) (Shalom *et al*, 2007). Mutations in T6SS I destroy the ability of the bacteria to form multinucleated giant cells, to escape from phagosomes, to survive and multiply inside macrophage cell lines and to remove the cytotoxicity of macrophages. The T6SS I mutants are attenuated in hamster models, which indicates that the T6SS I is essential for virulence and plays an important role in the intracellular lifestyle of *B. pseudomallei* (Burtnick *et al*, 2011).

1.1.5.9 The intracellular lifestyle of *B. pseudomallei*

B. pseudomallei has several mechanisms to escape macrophage killing and evade host immunity. It is able to survive and multiply within neutrophils and monocytes *in vitro* (Jones *et al*, 1996). These mechanisms include resistance to human defensin proteins (small cysteine-rich cationic proteins found in immune system cells to assist in killing phagocytosed bacteria) when exposed *in vitro* and the abolition of DNA and protein synthesis in host cells by secreting at least one uncategorized exotoxin (Jones *et al*, 1996). Moreover, *B. pseudomallei* is able to generate intracellular motility where the bacteria spread from one eukaryotic cell (host cell) to another by inducing continuous actin polymerization at one pole of the bacterial cell. Actin polymerization mobility allows cell to cell bacterial spread (Wiersinga *et al*, 2006). Also it is essential in the formation of multinucleated giant cells, although the mechanism by which this is produced is still obscure. However it has been observed in infected tissues from melioidosis patients, that *B. pseudomallei* is able to form multinucleated giant cells in both phagocytic and non-phagocytic cell (Kespichayawattana *et al*, 2000).

1.2 Methionine sulfoxide reductases, members of GAF domain family

1.2.1 The GAF family

The GAF domain family is one of the largest families of small-molecule binding domains that exists in nature and is widespread in all kingdoms of life (Martinez *et al*, 2002b). In mammals GAF domains are mostly present in cyclic nucleotide phosphodiesterases (PDEs), which are critical cellular enzymes controlling cGMP and cAMP second-messenger levels (Martinez *et al*, 2005). There are about 2000 GAF domain-containing proteins in which GAF domains have been shown to play several roles including the binding of small-molecules, protein-protein interactions (including dimerization) (Heikaus *et al*, 2009). The name GAF was given after the first three GAF domain-containing protein families were identified: the mammalian cyclic GMP-dependent phosphodiesterases (PDEs), *Anabaena* Adenylyl cyclase and *E.coli* transcription factor FhIA (Aravind & Ponting, 1997). However, the vast majority of GAF domains have not been studied, so their functions and ligand binding properties remain unclear (Heikaus *et al*, 2009).

1.2.1.1 GAF family history

The GAF domain was initially described by Aravind and Ponting in 1997 using the position-specific iterative BLAST method (Aravind & Ponting, 1997). The GAF domain has been described as one of the largest and most widely distributed families of small-ligand binding regulatory domains, although confirmed proofs of ligand binding remain insufficient, as few examples of GAF binding ligand are available (Zoraghi *et al*, 2004). The first demonstration GAF binding ligand was the binding of cGMP to mammalian PDE5 (Francis *et al*, 1980). Progressively more evidence has emerged. About 2000 proteins have been predicted to contain GAF domains but in humans they are found only in cyclic nucleotide phosphodiesterases (PDEs). Eleven different PDEs have been determined in mammalia. The PDEs are distinguished by a C-terminal conserved catalytic domain and one or more N-terminal regulatory domains, some of which bind

cGMP (Martinez et al, 2002b). Five PDE families have two regulatory domains in the N-terminus. Later on, by using bioinformatics techniques, it was shown that this regulatory domain exists in many other proteins, many of which do not bind cGMP; this regulatory domain was then renamed as a GAF domain (Aravind & Ponting, 1997). The first crystal structure of a GAF domain was identified by Ho and his colleagues in 2000. This GAF domain was seen in the structure of yeast YKG9 but no ligand for YKG9 has been reported (Ho *et al*, 2000). Then the first X-ray crystal structure of PDE GAF domains, with bound cGMP was determined in 2002; this structure revealed the dimerization role of the GAF domain where two GAF domains were found in PDE structure, and both have very similar folds to those seen in the single GAF domain of YKG9 (Martinez et al, 2002b).

1.2.1.2 GAF domain structure

All recorded structures of cyclic nucleotide-binding GAF domains indicate that the GAF domain consists of a core of a six-stranded antiparallel β sheets with the strand order 3-2-1-6-5-4. The N-and C-terminal portions of the sequence form one outer layer of the structure, composing of three helices: $\alpha 1$, $\alpha 2$ and $\alpha 4$ where the opposite outer layer of the structure is a mixture of loops and a short α -helix, $\alpha 3$, which is often packed against the domain and may take part in domain dimerization (Figure 1.4) (Heikaus *et al*, 2009; Ho *et al*, 2000). The first reported atomic resolution structure of the GAF domain was the crystal structure of the dimeric YKG9 (PDB code), a yeast protein of unknown function. Although no bound ligand was found in this structure, the overall topology of the domain suggests a cGMP binding role (Ho *et al*, 2000). In YKG9 the N-terminus extends 41 residues before the N-terminal end of the GAF domain as described by Aravind and Ponting (Aravind & Ponting, 1997). These 41 residues involve the first two helices, $\alpha 1$ (residues 8-14) and $\alpha 2$ (residues 20-38) where these helices form an integral portion of the α -helical layer, indicating that the stable folding unit for GAF domain corresponds to residues 8-179 of YKG9 (Ho *et al*, 2000). YKG9 dimerizes and most of the dimer interface is made up of residues of the N-terminal tail, $\alpha 3$, $\beta 2$, $\beta 3$ and the $\beta 2$ - $\beta 3$ loop (Figure 1.4). CYS91 and CYS125 form a disulfide bond in YKG9, this disulfide bond reduces the conformational freedom of the $\beta 2$ - $\beta 3$

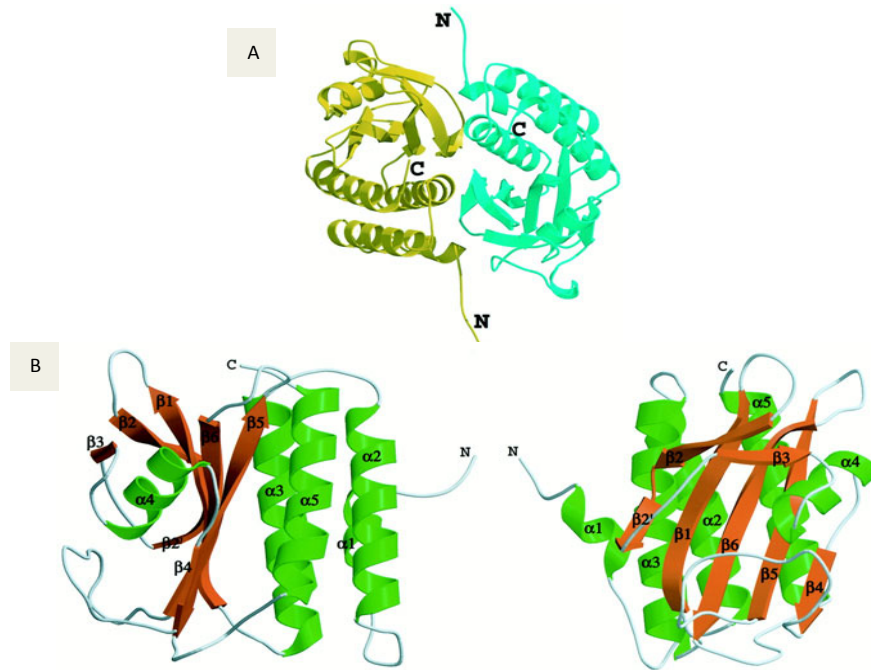


Figure 1.4: The structure of the GAF domain of yeast protein PDB code YKG9. (A) Overall structure of YKG9 dimer. (B) YKG9 monomer in two different views related by a 90° rotation about the y-axis. Adapted from (Ho *et al*, 2000).

loop and β 3- β 4 loop containing CYS91 and CYS125. It has been suggested that the disulfide linkage provides stability to the YKG9 crystal, since crystals grown in high concentration of reducing agent (dithiothreitol DTT) are small and diffract poorly (Ho *et al*, 2000). The first structure of any PDE GAF domain was the 2.9 Å resolution crystal structure of PDE2A tandem GAF domain. The structure shows a parallel homodimer in which GAF A contains no ligand but forms a dimerization link with a second GAF A, while GAF B does not contribute to the dimerization interface but binds cGMP in a deeply buried pocket (Figure 1.5) (Martinez *et al*, 2002a).

1.2.1.3 GAF domain functions

A subfamily of GAF domains has emerged as cyclic nucleotide (cNMP)-binding domains that allosterically control the catalytic activity of cyclic nucleotide phosphodiesterases (PDEs). Five PDEs (PDE2, 5, 6, 10 and 11) contain two N-terminal GAF domains, the first N-terminal domain is labeled as GAF A and the second C-terminal as GAF B (Heikaus *et al*, 2009). So far, only one GAF domain in each PDE monomer has been shown to bind cyclic nucleotide. It has been found that the GAF A domain of PDE5, 6, and 11 and the GAF B domain of PDE2 bind cGMP, while cAMP binds to the GAF B of PDE10 (Heikaus *et al*, 2009). PDEs regulate the cellular concentration level of the cyclic nucleotides cGMP and cAMP, both of which are essential second messengers in many signal transduction pathways (Martinez *et al*, 2002b). Cyclic GMP binds immediately to ion channels, activates protein kinase G (PKG), and controls the activity of phosphodiesterases. Cyclic AMP activates protein kinase A (PKA), which like PKG, phosphorylates a number of protein substrates. Adenylyl and guanylyl cyclases catalyze the synthesis of cyclic nucleotides, where 3,5-cyclic nucleotide phosphodiesterases (PDEs) hydrolyze these second messengers to the inactive 5-monophosphates (Martinez *et al*, 2002b). Because of their critical role in many disease-related pathways, PDEs are valuable drug targets. For example Vigra, Cialis and Levitra target PDE5 as it plays role in the male erectile pathway (Heikaus *et al*, 2009). However, in 2007 a new GAF domain function appeared when Lin and colleagues found that the *E. coli* free methionine-(*R*)-sulfoxide reductase (fRMs_r) was the first GAF domain family member that shows enzymatic activity, where the other GAF

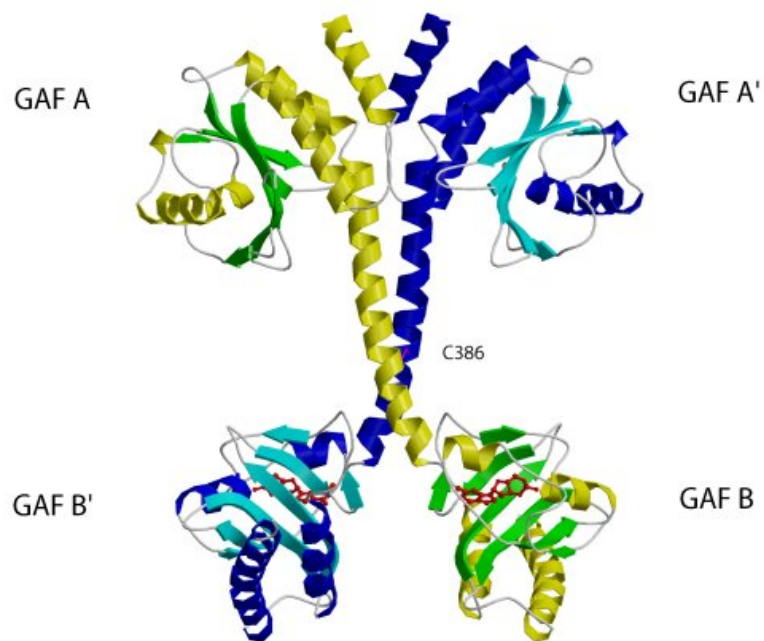


Figure 1.5: The structure of PDE2A the regulatory segment of mouse (PDB code 1MCO). Each PDE2A subunit contains a GAF A and a GAF B domain. The GAF A domain and seven turns of the connecting helices form a dimer interface. The two GAF B domains are far apart and contain the cGMP-binding sites. Adapted from (Martinez et al, 2002b)

domains bind cyclic nucleotides, chromophores and many other ligands for signal potentiation (Lin *et al*, 2007).

1.2.2 Methionine sulfoxide reductases

1.2.2.1 Reactive oxygen species (ROS) and oxidative stress

Reactive oxygen species (ROS) or free radicals are chemically reactive molecules containing oxygen, such as the superoxide anion and hydrogen peroxide. ROS are generated as part of the normal aerobic cellular existence and have essential roles in cell signaling and homeostasis but during times of environmental stress, ROS levels can increase dramatically, and this stimulates the formation of many other molecules able to cause oxidative stress in cells (Hoshi & Heinemann, 2001). Several non-enzymatic and enzymatic procedures are used to maintain the overall balance between ROS production and elimination (Finkel, 2000). ROS, specifically in the presence of cofactors such as certain metal ions (Fe^{2+} , Cu^{2+} , Co^{2+} and Mn^{2+}), are able to oxidatively damage and modify many cellular components such as DNA, proteins and lipids. This damage is likely to contribute to ageing, age-associated changes and age-related degenerative diseases like Alzheimer's and Parkinson's diseases (Hoshi & Heinemann, 2001; Stadtman, 1992). The oxidative stress induced by these reactive oxygen agents might cause DNA damage which may lead to cell death, abnormal cell growth and mutations, which in turn may result in cancer (Lee *et al*, 2009). It can lead also to lipid peroxidation which results from oxidation of cholesterol and fatty acids and may change membrane rigidity, and permeability, and may disrupt membrane networks and generate toxic products such as malondialdehyde (Nair *et al*, 2007). Oxidized proteins may become dysfunctional due to structural modification and catalytic failure (Lee *et al*, 2009). Amino acids in proteins have a high tendency to be oxidized by the ROS and this oxidative change affects the backbone as well as the amino acid side chains. Side chain oxidation can modify the overall chemistry of the amino acids which spontaneously changes protein function (Stadtman & Berlett, 1998). The sulfur-containing amino acids, cysteine and methionine, are the most sensitive residues to ROS-mediated oxidation, but also the side chains of lysine, histidine,

arginine, proline, tryptophan and tyrosine are prone to be oxidized (Stadtman & Berlett, 1998). Moreover the thiol-containing cysteine oxidation is regularly stimulated by the presence of small amounts of metal ions such as Fe^{2+} , Cu^{2+} , Co^{2+} and Mn^{2+} producing several products like sulfenic ions, disulfides and sulfonic ions. Disulfide formation is the most common result of cysteine oxidation, but this reaction can be easily reversed to thiol by using reducing agents such as glutathione *in vivo* or dithiothreitol (DTT) *in vitro* (Finkel, 2000).

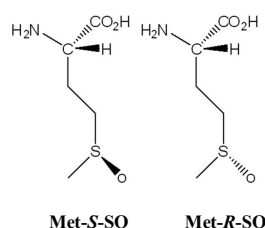
1.2.2.2 Oxidation of methionine

Methionine is a sulfur-containing amino acid susceptible to oxidization by a range of different ROS, such as $\text{O}_2^{\cdot-}$, H_2O_2 , peroxynitrite (ONOO^-) or $\cdot\text{OH}$ to methionine sulfoxide (Met-SO). The formation of methionine sulfoxide has been recorded in native proteins and reveals that methionine side chain oxidation is a physiologically consistent phenomenon. Methionine oxidation to methionine sulfoxide (Met-SO) or even further oxidation to methionine sulfone (Met-SO_2) is expected to induce significant changes in the methionine side chains physical properties, leading to protein function alteration (Hoshi & Heinemann, 2001). A change in hydrophobicity is one aspect of these changes where the side chain of the normal methionine is long, flexible, and non-polar but the side chain of methionine sulfoxide with the additional oxygen atom is stiffer and more polar than the methionine side chain. The specific oxidation and reduction of methionine residues in proteins may cause serious effects in protein function and cell physiology. Methionine oxidation is considered as one of the essential factors that contribute in physiological dysfunctions to the long-term age-associated changes and degenerative diseases. Moreover methionine oxidation is expected to play a role in physiological dysfunctions if there is acute local generation of excess ROS (Hoshi & Heinemann, 2001). Several examples of protein damage by methionine oxidation leading to critical clinical issues have been described, like reperfusion injury resulting from ischaemic episodes, which can take place in almost all organs but notably in brain and cardiac tissues. Reperfusion occurs in the condition of ample production of ROS and cell damage (Chan, 1996). The proteins damaged by oxidation play a significant role in inflammatory processes (Winrow *et al*, 1993).

The level of methionine oxidized protein increases with age (Stadtman, 1992) in different systems especially the human brain, human eye lens and human erythrocytes. Various degenerative diseases have been linked to methionine oxidation proteins such as Parkinson's disease, Alzheimer's disease and eye lens cataracts (Hoshi & Heinemann, 2001). Also animal models show that old animals are more sensitive to oxidative stress and the life period of animals corresponds to the amount of oxidized protein (Yu *et al*, 1998). However, there are many proteins that have methionine side chains and oxidation of these residues does not seem to damage function. Accordingly, it has been suggested that the cyclic oxidation/reduction of these methionine residues is regulated by an anti oxidative system (Lee *et al*, 2009).

1.2.2.3 Methionine sulfoxide reductase (Msr) family

Oxidation of methionine produces a diastereomeric mixture of methionine *S*-sulfoxide (Met-*S*-SO) and methionine *R*-sulfoxide (Met-*R*-SO) (Weissbach *et al*, 2005):



Methionine sulfoxide (Met-SO) has been involved in a variety of diseases and in the aging process. Several types of methionine sulfoxide reductase (Msr) mediate the reduction of Met-SO. The first Msr enzyme called MsrA was identified in 1978, and it was found that this enzyme is able to rescue the function of oxidized ribosomal protein L21 in *E. coli* by reducing the oxidized state of the protein (Caldwell *et al*, 1978). After the discovery of the first Msr enzyme, many Msrs have been found in a wide range of organisms, and currently these enzymes can be classified into three types (Lee *et al*, 2009):

1. MsrA:

MsrA catalyzes the reduction of free and protein-based methionine-*S*-sulfoxide

(Met-*S*-SO) (Lee *et al*, 2009). It is a small cytosolic enzyme that exists in a wide range of organisms from bacteria to plants and mammals, including humans. The amino acid sequence is well conserved among various species (Hoshi & Heinemann, 2001). In mammals MsrA is the only recognized enzyme able to reduce methionine-*S*-sulfoxide (Moskovitz *et al*, 2000). It has been found that MsrA is expressed in different tissues in rat such as liver, kidney, heart and brain (Moskovitz *et al*, 1996). In adult humans it is highly expressed in liver and kidney and this corresponds with the theory that MsrA plays a role as an anti-oxidant enzyme. In addition MsrA has been found in human heart and brain, and the brain also has a high level of overall MsrA expression (Kuschel *et al*, 1999) but it is highest in the cerebellum which suggests that methionine oxidation may play a role in neural function (Hoshi & Heinemann, 2001). At the subcellular level MsrA is found in the cytosol, mitochondria and nucleus, however the molecular mechanism and regulation remains unclear (Kim & Gladyshev, 2005b).

2. MsrB:

The second Msr named MsrB is responsible for the reduction of protein-based methionine-*R*-sulfoxide (Met-*R*-SO) and also it may catalyze the reduction of free methionine-*R*-sulfoxide, but with low efficiency (Lee *et al*, 2009). For example, the reduction capability of *E.coli* MsrB on the free methionine-*R*-sulfoxide is 1000 fold lower than the catalytic ability of *E.coli* MsrA for the reduction of free methionine-*S*-sulfoxide (Grimaud *et al*, 2001). Moreover in humans MsrB is not efficient at reversing the free methionine-*R*-sulfoxide to free methionine (Lee *et al*, 2009). In mammals, MsrBs have been classified into three types: MsrB1, MsrB2 and MsrB3 according to their catalytic mechanism. All three types have Zinc (Zn), coordinated by two CxxC motifs (two cysteines separated by two residues) which stabilizes the MsrB structure (Kryukov *et al*, 2002). Among three mammalian methionine-*R*-sulfoxide reductases (MsrBs), MsrB1 is a Selenocysteine-containing protein, whereas MsrB2 and MsrB3 contain CYS in the active site (Kim & Gladyshev, 2005a). Mammalian MsrB1 occurs in the cytosol and nucleus, MsrB2 in the mitochondria and MsrB3 in endoplasmic reticulum and mitochondria (Lee *et al*, 2009).

3. fRMsR:

The third type of Msr is the free methionine-*R*-sulfoxide reductase (fRMsR). This type was discovered in 2007 when Lin and his colleagues purified the enzyme from the MsrA and MsrB strain and determined the identity of (fRMsR) through proteomic analysis. The fRMsR domain was identified as a member of the large GAF domain family, which were initially known as cGMP-binding phosphodiesterases. Sequence analysis of fRMsR shows that the domain is conserved among unicellular organisms, like bacteria and yeast, but it is absent in multicellular organisms including mammals (Lin *et al*, 2007).

The three Msr types maintain proteins function by catalyzing the reduction of free methionine sulfoxide and methionine sulfoxide-containing proteins. Methionine sulfoxides and Msrs may also contribute in sensing changed redox (reduction-oxidation reactions) status and regulating certain proteins (Lee *et al*, 2009). However, the three types of Msr have different catalytic features: The mammalian MsrA and fRMsR have three conserved cysteines involved in the reaction as catalytic and resolving cysteines. The catalytic cysteine attacks the sulfur of methionine sulfoxide generating a cysteine sulfenic acid intermediate (CYS-SOH) with release of methionine. The generated thiol (catalytic cysteine bonded sulfenic acid) forms a disulfide bond with the resolving cysteine, which is finally reduced by thioredoxin (Figure 1.7). The difference between MsrA and fRMsR includes the binding site volume, where the fRMsR has a narrow binding pocket, which restricts access to bulky substrates. This difference corresponds to the substrate specificity of both types (Boschi-Muller *et al*, 2008; Lin *et al*, 2007; Lowther *et al*, 2000). MsrB enzymes have various mechanisms. MsrB1 possesses one conserved cysteine in the N-terminal part and the catalytic selenocysteine in the C-terminal part. A selenic acid intermediate at the catalytic selenocysteine is produced when this catalytic residue attacks the sulfur of methionine-*S*-sulfoxide and then this intermediate rearranges into selenenylsulfide with the assistance of the resolving cysteine and it is finally reduced by thioredoxin (Kim & Gladyshev, 2005a). In comparison MsrB2 and MsrB3 have only one conserved cysteine, which suggests a different catalytic mechanism where the sulfur of methionine-*S*-sulfoxide is attacked, by the catalytic cysteine and the sulfenic acid intermediate produced and

directly reduced by thioredoxin without disulfide formation between catalytic and resolving cysteines (Kim & Gladyshev, 2005a).

1.2.2.4 Role of methionine sulfoxide reductases in bacterial pathogenicity

The methionine sulfoxide reductases (Msr) may be a determinant agent for bacterial pathogenicity. Several recorded studies show that these enzymes may play significant roles in bacterial virulence and they assume that these roles are related to the function of the methionine sulfoxide reductases as antioxidant systems (Ezraty *et al*, 2005). In one of these studies for example, they found that *Erwinia chrysanthemi*, which is a plant pathogen, attacks plants by producing a massive amount of plant cell degradation, can prevent the infection of the whole plant by destroying the MsrA in this bacterium. They found a significant difference in virulence of the wild type *Erwinia chrysanthemi* and the MsrA mutant *Erwinia chrysanthemi* (El Hassouni *et al*, 1999). Another example of restricting bacterial virulence by generating Msr mutants is in *Staphylococcus aureus*, a leading cause of hospital and community-acquired infection (Singh & Moskovitz, 2003). Also it has been shown that MsrA and MsrB mutants in *Neisseria gonorrhoeae* and *Neisseria meningitidis* are not able to cause the sexually transmitted infection Gonorrhoea and the meningitis respectively, while wild types of both pathogens are capable of producing the diseases (Shafer & Rest, 1989).

1.2.3 Examples of methionine sulfoxide reductases containing GAF domain

All the known methionine sulfoxide reductases containing GAF domains are of the free methionine-*R*-sulfoxide reductase fRMsr type (Lee *et al*, 2009). The genomic analysis of fRMsr showed that its existence is restricted to unicellular organisms, where multicellular organisms, including mammals lack this enzyme (Lee *et al*, 2009). So far, four crystal structures of fRMsr are available and they are deposited in the Protein Data Bank with the following codes: 1VHM, 3KSF, 3MMH and 3RFB (3RFB has been identified as an fRMsr in PDB but is still not published).

These enzymes are from *Escherichia coli*, *Staphylococcus aureus*, *Neisseria meningitidis* and *Streptococcus pneumonia*, respectively.

1. Free methionine-*R*-sulfoxide reductase from *Escherichia coli* (1VHM):

The crystal structure of *E. coli* fRMsR in complex with a MES buffer molecule has been solved by Badger and his colleagues (Badger *et al*, 2005) but no known function was recorded, this structure has the code 1VHM in Protein Data Bank (PDB). Two years later, the type three Msr, fRMsR was first named and identified by Lin and his colleagues (Lin *et al*, 2007). In this study they determined and characterized the *Escherichia coli* fRMsR by its extraction and purification from the MsrA⁻B⁻ strain and proteomic analysis. The fRMsR sequence is highly conserved across bacteria and yeast but is absent in multicellular organism including humans. The sequence database shows many homologous sequences including many hypothetical proteins of unknown function (Lin *et al*, 2007). The recombinant fRMsR DNA was created and the histidine tagged fRMsR protein was expressed with a molecular weight of ~24 kDa. The reducing system and substrate specificity of *E. coli* fRMsR were examined by evaluating the reductase activity using different test conditions, containing substrates of different sizes and isomers of Met-SO, and the reductase activity was estimated by monitoring the decrease in NADPH absorbance at 340 nm. fRMsR exhibited the highest Met-SO reductase activity when all components of the NADPH-TrxR-Trx system were present in the reaction. The enzyme shows precise substrate specificity to free Met-*R*-SO, and it was not able to catalyze the reduction of Met-*S*-SO, Met sulfone, dimethyl sulfoxide or the Met-SO when binding the synthetic peptide NH₂-Pro-Thr-Ser-Met-(*RS*)-O-Glu-His-Val-NH₂. In contrast they found that the control proteins, bovine MsrA and *Neisseria gonorrhoeae* MsrB are able to reduce both free and peptide bound forms of Met-*S*-SO and Met-*R*-SO, respectively (Lin *et al*, 2007). The 2.1 Å resolution crystal structure of *E. coli* fRMsR in complex with a MES buffer molecule consists of six antiparallel β-strands, four α-helices and two loops. The structure shows that the MES sulfonic acid group is close to CYS94 and the disulfide bond between CYS84 and CYS118 and this is similar to the arrangement of the active site residues in MsrA (Figure 1.6) (Lin *et al*, 2007). Lin et al's search

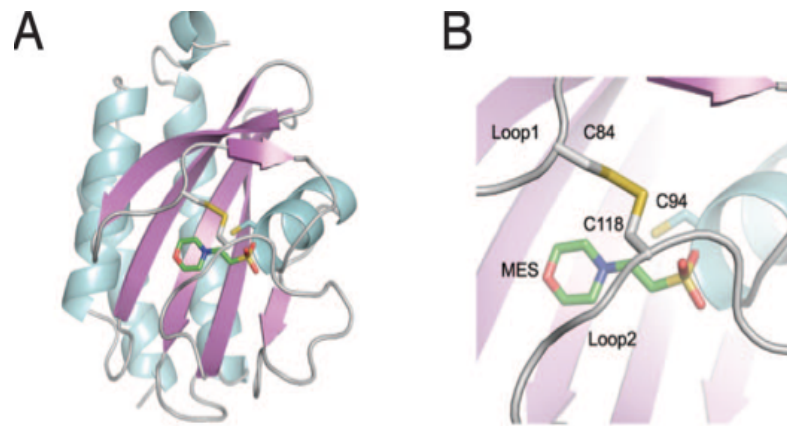


Figure 1.6: *E. coli* fRMsR structure (PDB code 1VHM). (A) Overall structure shows the MES molecule in stick rendering and located in the active site. (B) Close up in cartoon representation view of the active site shows the MES molecule close to the three conserved CYS residues: CYS 84, CYS 94 and CYS118. The disulfide bond is shown between CYS 84 and CYS 118 linking loop1 and loop2 together. Adapted from (Lin *et al.*, 2007)

of the DALI database with the *E. coli* fRMsR structure showed a high homology to the GAF domain family suggesting that the fRMsRs use three cysteine residues for catalysis and formation of a disulfide bond to enclose a small active site cavity. The *E. coli* fRMsR has been considered as the first a GAF domain family member to exhibit enzymatic activity, as the previous only known function of GAF domain was binding cyclic nucleotides for signal potentiation. This suggests that the free Met-*R*-SO may act as signaling molecule in response to oxidative stress (Lin *et al*, 2007).

2. Free methionine-*R*-sulfoxide reductase from *Staphylococcus aureus* (3KSF)

Crystal structures of the reduced, substrate-bound and oxidized forms of fRMsR from *Staphylococcus aureus* have been obtained by Bong and his colleagues (Bong *et al*, 2010). *Staphylococcus aureus* is a leading cause of hospital-and community-acquired infections (Singh & Moskovitz, 2003). The *S. aureus* fRMsR sequence reveals that the enzyme like the *E. coli* fRMsR, contains three conserved cysteine residues (CYS68, CYS78 and CYS102) for catalyzing the free methionine-*R*-sulfoxide reduction. Also the *S. aureus* fRMsR has a narrow and small active site and these enclosed cavities in all fRMsR enzymes confirm the obvious substrate specificity for free methionine-*R*- sulfoxide in comparison with the peptide based methionine sulfoxide which requires a wider active site. It had been suggested in previous studies that the catalytic cysteine in *Staphylococcus aureus* fRMsR is CYS 78 and the disulfide bond between CYS68 and CYS102 (Lin *et al*, 2007). But in the Bong study the biochemical analysis suggests CYS102 as a catalytic cysteine but also it forms a disulfide bond with CYS68 while, the CYS78 that was previously thought to be a catalytic cysteine, is not essential for catalytic function (Bong *et al*, 2010). In this study the biochemical analysis was achieved by using the wild type *Staphylococcus aureus* fRMsR and single and double mutants where the three conserved cysteines were replaced with serine residues. To decide what the function of each cysteine residues (CYS68, CYS78 and CYS102) was, the following single and double mutants were created: (C68S, C78S, C102S, C68S/C78S and C68S/C102S) and they compared the Trx-dependent activities of these mutants with the wild type. Interestingly, they found that the C78S mutant

retains 75% of enzyme activity in comparison with the wild type, although the CYS78 is thought to be the catalytic residue in *Escherichia coli* fRMs_r. The mutant C68S/C78S kept 22% of enzyme activity but in contrast, C102S had no catalytic activity and C68S/C102S had no activity either. The catalytic activity of C68S was 32% of the wild type. Accordingly, Bong *et al* suggested CYS102 is the catalytic cysteine, CYS68 may act as the resolving cysteine that forms a disulfide bond with CYS102, and CYS78 is not critical for catalytic activity. So the suggested catalytic mechanism of *Staphylococcus aureus* fRMs_r is: CYS102 attacks free methionine-*R*-sulfoxide leading to the formation of the thiol (Cys-sulfonic acid intermediate) and releasing of methionine. CYS68 interacts with the Cys-sulfonic acid to form a disulfide bond. Finally the disulfide bond is reduced by a reductant (Trx *in vivo* or DTT *in vitro*) (Figure 1.7) (Bong *et al*, 2010). Structural comparisons between the reduced substrate-bound and oxidized forms of *Staphylococcus aureus* fRMs_r suggests that conformational changes appear in the active site during catalysis (Figure 1.8) (Bong *et al*, 2010).

3. Free methionine-*R*-sulfoxide reductase from *Neisseria meningitidis* (3MMH)

Interestingly, another fRMs_r structure from *Neisseria meningitidis* presents conflicting results. The crystal structure of oxidized *Neisseria meningitidis* fRMs_r was solved at 1.25 Å resolution in complex with free methionine-*R*-sulfoxide (Gruez *et al*, 2010). Like all the fRMs_rs, *Neisseria meningitidis* fRMs_r contains three conserved cysteines (CYS118, CYS94 and CYS84) but also had an additional cysteine residue in position 136. Biochemical and kinetic data from the wild type and mutants in which the cysteines were replaced by alanine was performed to determine the catalytic mechanism. Single, double and triple mutants were made and the Trx-dependent activities of these mutants were compared with the wild type. They found that all mutants remained active except the mutant C118A which indicates that the catalytic cysteine is CYS118. This result is consistent with *Staphylococcus aureus* fRMs_r where the CYS118 corresponds to CYS102 in *Staphylococcus aureus* (Gruez *et al*, 2010). But a significant difference was noticed between the crystal structure of the substrate-bound form of *Staphylococcus aureus* fRMs_r and the structure of oxidized *Neisseria meningitidis* fRMs_r when binding

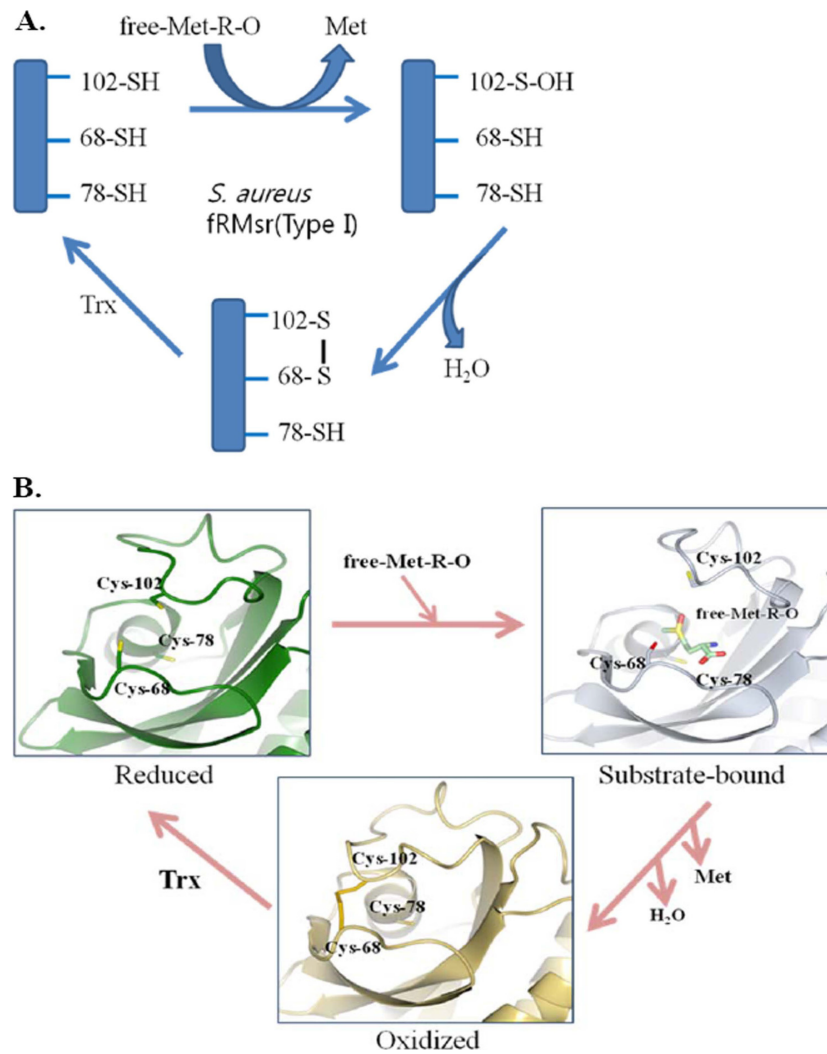


Figure 1.7: *Staphylococcus aureus* fRMsR (PDB code 3KSF) reduction activity. (A) Suggested catalytic mechanism of *Staphylococcus aureus* fRMsR. (B) Structural comparison shows conformational changes in the active site. Reduced fRMsR exhibits an open conformation in the active site and after binding of substrate the enzyme changes to substrate-bound to closed conformation. Oxidized fRMsR displays a more closed conformation by the formation of the disulfide bond. Adapted from (Bong *et al*, 2010).

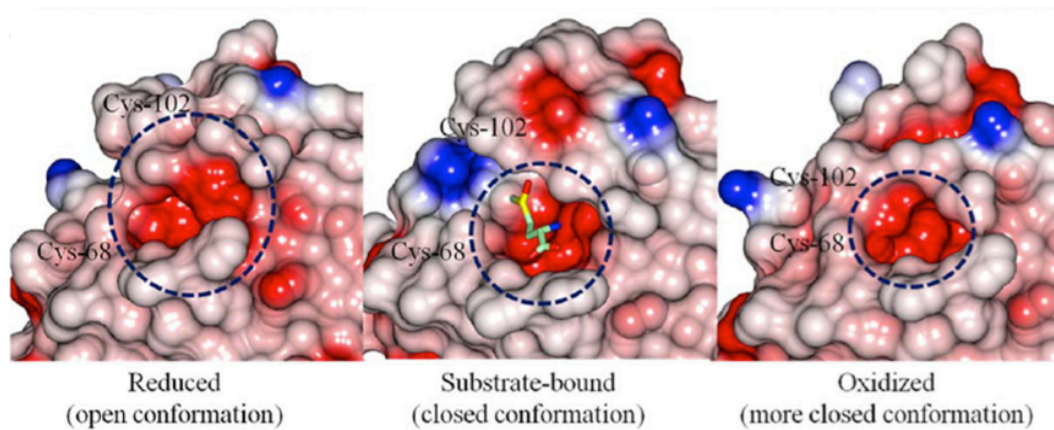


Figure 1.8: Conformational changes of fRMsr_{red}, fRMsr_{sub}, and fRMsr_{ox} forms of *Staphylococcus aureus* fRMsr (PDB code 3KSF). The active site is shown with electrostatic surface models. Adapted from (Bong *et al*, 2010).

substrate. The difference is in the orientation of the position of substrate binding where the sulfoxide moiety in the *Staphylococcus aureus* fRMsR replaces the position of the carboxylate group in *Neisseria meningitidis* fRMsR and vice versa (Bong *et al*, 2010; Gruez *et al*, 2010). However the *Staphylococcus aureus* fRMsR substrate bound structure was obtained using a mutant of CYS68 (corresponding to CYS84 in *Neisseria meningitidis* fRMsR), which is the resolving cysteine that forms the disulfide bond with the catalytic cysteine CYS102 (CYS118 in *Neisseria meningitidis* fRMsR) (Bong *et al*, 2010). While the oxidized- substrate bound complex of *Neisseria meningitidis* fRMsR form a disulfide bond between CYS118 and CYS84 (Gruez *et al*, 2010). The structure of *Neisseria meningitidis* fRMsR (Figure 1.9) shares the overall topology of GAF domains. The structure of a subunit is arranged around a central twisted antiparallel β -sheet composed of six strands (β 3- β 2- β 1- β 6- β 5- β 4) bounded on one side by a three- α -helix bundle comprising the helices α 1, α 2 and α 4, and on the opposite side by a unit consists of loop- β 3- α 3-loop (Gruez *et al*, 2010). Loop- β 3- α 3-loop motif links the overall β -sheet and the loop between β 4 and β 5 strands functions as a capping flap that shield the active site.

1.2.4 A new free methionine-(*R*)-sulfoxide reductase from *Burkholderia pseudomallei* (BPSL2418) belongs to the GAF domain family

The rise of drug resistance in pathogenic microorganisms represents a major challenge to human health and life. *Burkholderia pseudomallei* is one of the high resistance organism to antibiotics including the currently used drug, ceftazidime (Thibault *et al*, 2004). The growing number of microbial genome sequencing projects and the improvement of bioinformatics tools facilitate *in silico* analysis of gene sequence information (Kaminski, 2000). This computer-aided facility is useful in the identification, validation, and selection of the potential genes leading to further experimental analysis and characterization, which could also be potential drug candidates. Recently *in silico* analysis of the *Burkholderia pseudomallei* genome sequence has identified 312 essential genes, which might be drug targets for melioidosis. These genes encode essential proteins that play roles in the

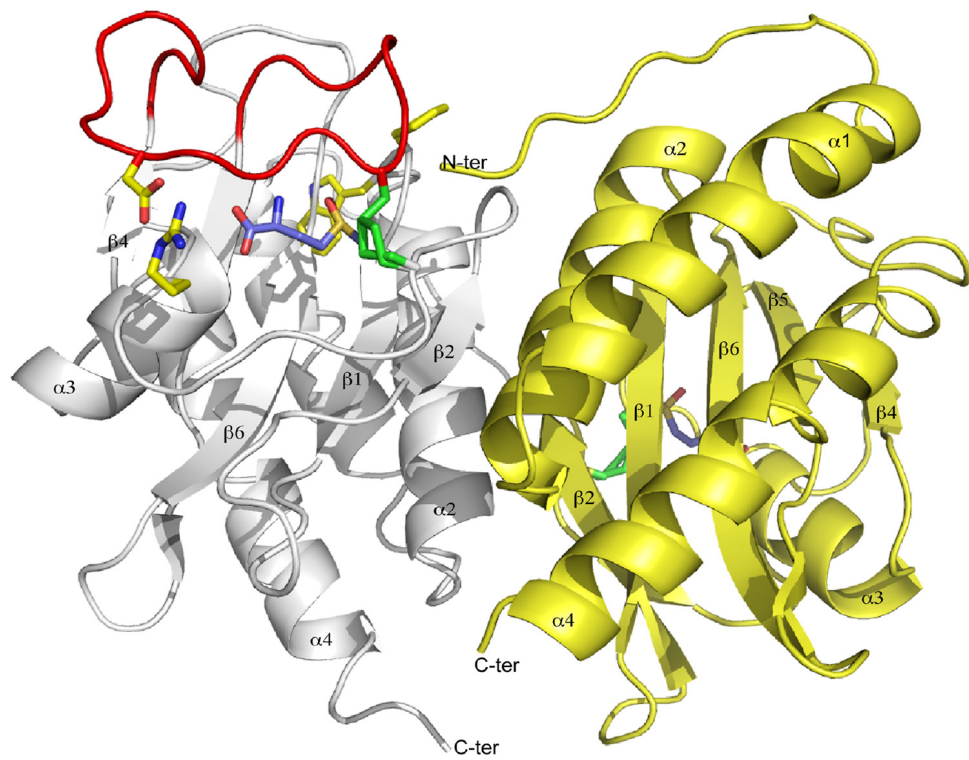


Figure 1.9: Cartoon ribbon of oxidized- binding substrate complex of *Neisseria meningitidis* fRMsR (PDB code 3MMH). The two subunits are in white and yellow. The flap of the gray subunit is in red. The Met-R-SO is represented in a stick representation with the carbon atoms in purple, the oxygen atoms in red, the nitrogen in blue and sulfur atoms in yellow. Adapted from (Gruetz *et al*, 2010).

survival of *B. pseudomallei* including outer and inner membrane proteins, regulators, proteins involved in pathogenicity, adaptation, chaperones as well as proteins involved in degradation of small molecules and macromolecules, energy metabolism, information transfer, central/ intermediate/ miscellaneous metabolism pathways and some conserved hypothetical proteins of unknown function (Chong *et al*, 2006). Also a proteomic analysis between pathogenic strains of *Burkholderia* determined a list of possible candidates that might play roles in *B. pseudomallei* pathogenicity. BPSL2418 is one of these putative targets in *B. pseudomallei* and it has been identified as a conserved hypothetical protein of unknown function (Chong *et al*, 2006; Wongtrakoongate *et al*, 2007). Sequence analysis shows that BPSL2418 gene encodes a protein containing a GAF conserved domain (Figure 1.10). A blast search was conducted to find potential homologs in other organisms of known function or structure, and shows that BPSL2418 shares high homology with the GAF domain fRMsrs enzymes. High sequence identity has been found between BPSL2418 and the fRMsrs of known structure, 1VHM, 3KSF, 3MMH and 3RFB, which suggest that BPSL2418 might act as methionine sulfoxide reductase (Figure 1.11). Amino acid sequence alignment between the fRMsrs including BPSL2418 shows that the three cysteines that are thought to catalyze the reduction are fully conserved. BPSL 2418 contains the three conserved cysteines (CYS75, CYS85 and CYS109) in the active site and an additional cysteine residue in position 154 (Figure 1.12).

In the Protein Data Bank a crystal structure of peptide methionine-*R*-sulfoxide reductase in *Burkholderia pseudomallei* (PDB code 3CEZ) is available, but up to date this work is not published. This protein was represented as an MsrB family member and it has similar structure features of the MsrBs, which are composed of 8 β strands and two N-terminal α helices (Figure 1.13). The 3CEZ protein structure consists of one monomer and has an acetic acid binding ligand.

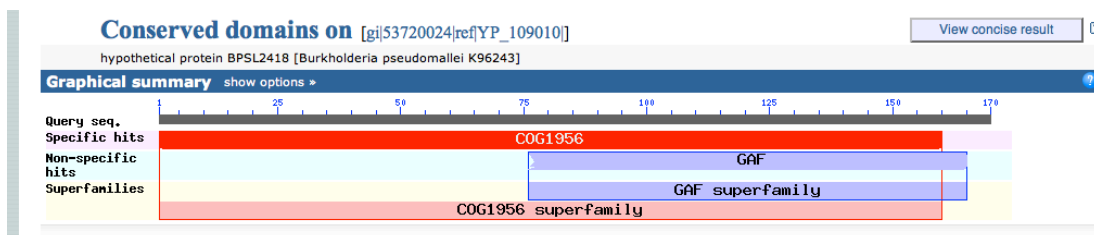


Figure 1.10: BPSL2418 is a GAF domain-containing protein. Available from the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>).

1VHM	3KSF	3RFB	3MMH
Identity: 51%	Identity: 49%	Identity: 43%	Identity: 49%
Similarity: 65%	Similarity: 68%	Similarity: 65%	Similarity: 60%
Method of identification: X-ray	Method of identification: X-ray	Method of identification: X-ray	Method of identification: X-ray
Organism: <i>Escherichia coli</i>	Organism: <i>Staphylococcus aureus</i>	Organism: <i>Streptococcus pneumoniae</i>	Organism: <i>Neisseria meningitidis</i>
Function: free Met-R-O reductase	Function: free Met-R-O reductase	Function: free Met-R-O reductase	Function: free Met-R-O reductase
Contains GAF domain: YES	Contains GAF domain: YES	Contains GAF domain: YES	Contains GAF domain: YES
Ligand: MES buffer,	Ligand: DI(HYDROXYETHYL)ETHER	Ligand: Free methionine sulfoxide	Ligand: Free methionine sulfoxide

Figure 1.11: The percentages of sequence identity and similarity between the hypothetical protein from *Burkholderia pseudomallei* BPSL2418 and the GAF domain fRMsrs enzymes: *Escherichia coli* fRMsr, *Staphylococcus aureus* fRMsr, *Neisseria meningitidis* fRMsr and *Streptococcus pneumoniae* fRMsr which have the Protein Data Bank codes 1VHM, 3KSF, 3MMH and 3RFB, respectively.

```

3MMH      MH-ALHFSASDKAALYREVLPOIESVVADETDWVANLANTA AVLKEAF-GWFWVGFYLVLD
BPSL2418  MFALSEALPSSKPALYETLAAQARALVETETDIVANAANFASLVYHSLDGLNWAGFYFFD
1VHM      -----MNKTEFYADLNRFNALMAGETSFLATLANTSALLYERLTDINWAGFYLLE
3KSF      -----MTTINPTNYTLLKKQAASLIEDEHHMIAILSNMSALLNDQINWVGFYLLE
           : * : : ::: * :* :* :::: . : *.*.*.:.

3MMH      TRSDELVLAPFQGPLACTRIPFGRGVCGQAWAKGGTVVVGDVDAHDPDHIACSSLSRSEIV
BPSL2418  G--RELVVGPFQGKPACVRIPLGKGVCGTAAQTRGTQVVHDVHAFAGHIACDSASQSEIV
1VHM      D--DTLVLGPFQGKIACVRIPVGRGVCGTAVARNQVQRIEDVHVFDGHIACDAASNSEIV
3KSF      Q--NELILGPFQGHPACVHIPIGKGVCGTAVSERRTQVVADVHQFKGHIACDANSKSEIV
           *:.*.*.* * :.*.*.*.* * . : ** . . *.*.*.:*.*.*.*

3MMH      VPLFSD-GRCIGVLDADSEHLAQFDETDALYLGELAKILEKRFEASRQAV-----
BPSL2418  VPLVARDGALIGVWDVDSPLVGRFDAEDAIGMEALCRVFVEVAWENATRGE-----
1VHM      LPLVVK-NQIIGVLDIDSTVFGRFTDEDEQGLRQLVAQLEKVLATTDYKKFFASVAG
3KSF      VPIFKD-DKIIGVLDIDAPITDRFDDNDKEHLEAIVKIIEKQLA-----
           :*:. *.* * * : : * * : : : :


```

Figure 1.12: Amino acid Alignment of BPSL2418, *E. coli* fRMsr (1VHM), *Staphylococcus aureus* fRMsr (3KSF) and *Neisseria meningitidis* fRMsr (3MMH) proteins. (*) indicates a fully conserved residue. (:) indicates conservation between groups of strongly similar properties. (.) indicates conservation between groups of weakly similar properties. The three fully conserved cysteines are in red. Cysteines in green are non-conserved residue. The amino acid alignment was achieved using Clustal omega (www.ebi.ac.uk/Tools/msa/).

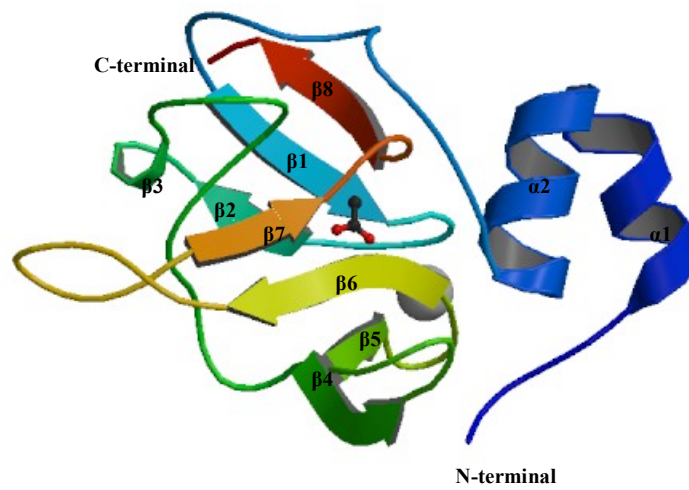


Figure 1.13: Cartoon representation of peptide methionine-*R*-sulfoxide reductase in *Burkholderia pseudomallei* (3CEZ) protein. The structure is composed of 8 β strands and two N-terminal α helices with an acetic acid binding ligand. Adapted from the Protein Data Bank.

1.3 Aims of the projects

Melioidosis has emerged as a lethal disease threat in the endemic areas and potentially all over the world. Unfortunately, the molecular mechanisms of virulence and pathogenicity of *Burkholderia pseudomallei* remain unknown. But genomics and proteomics research have selected putative targets in *Burkholderia pseudomallei* that might play roles in the virulence and pathogenicity of the organism. This project aims to determine the structure of one of the hypothetical unknown function target proteins in *Burkholderia pseudomallei*, BPSL2418. It has been predicted from the sequence analysis that BPSL2418 might function as a free methionine sulfoxide reductase and interestingly there is a link between the level of the methionine sulfoxide in pathogen tissues and its virulence (Ezraty *et al*, 2005). Currently few structures of free methionine sulfoxide reductase are known, and the more structures of this domain to be solved will help in understanding their role and the mechanism of these enzymes.

Therefore in this thesis I describe the cloning expression, purification and crystallization of BPSL2418 and the solution of its 3D structure using X-ray crystallography. Also we aimed to identify the substrate bound, MES bound, unbound and reduced forms of the enzyme in order to increase the understanding of the catalytic mechanism of free methionine sulfoxide reductase.

Second project: The C-terminal domains of Polycystin 1 (CT1) and Polycystin 2 (CT2)

1.4 Polycystin 1 and Polycystin 2

Polycystin 1 and polycystin 2 are human members of a conserved family of proteins that includes homologues in the mouse, the Fugu fish, the sea urchin and the worm *Caenorhabditis elegans* (Wilson, 2001). They are highly expressed on renal tubular epithelia (Wilson, 2004). Mutations in genes encoding polycystin 1 and polycystin 2 are associated with autosomal dominant polycystic kidney disease, the most common human genetic disease, with a prevalence of 1 in 800 of the world's population (Wilson, 2001). Although the function of these proteins is still unknown, many studies suggest that polycystin 1 is involved in cell-cell or cell-matrix interactions whereas polycystin 2 is thought to be a channel protein (Wilson, 2001). Polycystin 1 is a large membrane protein with a long extracellular N-terminal portion, 11 transmembrane domains and a 200 amino acid intracellular C-terminal portion, whereas polycystin 2 is predicted to be a membrane protein with six transmembrane-spanning domains and both the N- and the C-termini are predicted to be intracellular. A cytoplasmic interaction has been suggested between PC1 and PC2, but no structural study has proved the PC1-PC2 interaction (Wilson, 2001).

1.4.1 Autosomal Dominant Polycystic Kidney Disease (ADPKD):

Autosomal Dominant Polycystic Kidney Disease is the most common lethal inherited disorder that affects the kidneys. It occurs in 1 in 800 live births (Fogazzi, 1998). Two types of ADPKD are classified:

Type I is caused by mutations in the Polycystic Kidney Disease1 (PKD1) gene, which encodes the polycystin 1 protein. ADPKD type I accounts for 85 to 90 percent of patients.

Type II is caused by mutations in the Polycystic Kidney Disease 2 (PKD2) gene, which encodes Polycystin 2 protein, and accounts for 10-15% of all ADPKD cases (Reeders *et al*, 1988). ADPKD is a very widespread genetic disease, and parents with ADPKD have a 50 percent chance of passing the disease on to each of their children (Wilson, 2001). In addition to the autosomal dominant inherited disorder, there is also a recessive form called Autosomal Recessive Polycystic Kidney Disease (ARPKD). However ARPKD is much less frequent than ADPKD (Stawicki & Lombardo, 2008). One of the significant reasons of renal failure is the formation of fluid-filled cysts due to ADPKD. The fluid-filled cysts are formed by dysfunction in renal epithelial cells, and they cause a massive enlargement of both kidneys (Figure 1.14) (Chapin & Caplan, 2010). Polycystic kidney disease may be acquired in adult life as a consequence of aging, drugs and hormones, but most forms of PKD are hereditary (Wilson, 2004).

1.4.2 Polycystin 1

Polycystin 1 is a large membrane protein, which is subdivided into three parts or domains (extracellular, transmembrane and intracellular domains) (Figure 1.15).

1.4.2.1 Domain structure of Polycystin 1

Analysis of the amino acid sequence produced from the cloned full-length PKD1 gene using UniProt (Universal Protein Resource) predicted that polycystin 1 is a massive membrane protein of 4303 amino acids and molecular weight ≥ 462 kDa with a long extracellular N-terminal portion, 11 transmembrane helices and short intracellular C-terminal portion of ≥ 200 amino acids (Hughes *et al*, 1995) (Figure 1.15).

Extracellular Domain: The first > 3000 amino acids form the long extracellular portion that consists of multiple domains of various types such as: Leucine-rich, C-type lectin motif, Low-density lipoprotein A (LDL-A) domain, Immunoglobulin like Polycystic Kidney Disease (Ig-like PKD) domains and Receptor for Egg Jelly (REJ) domain (Wilson, 2001).

Transmembrane Domain: Hydropathy analysis of polycystin 1 shows that



Figure 1.14: Kidneys affected by Polycystic Kidney Disease.

Adapted from <http://pathologyproject.files.wordpress.com>

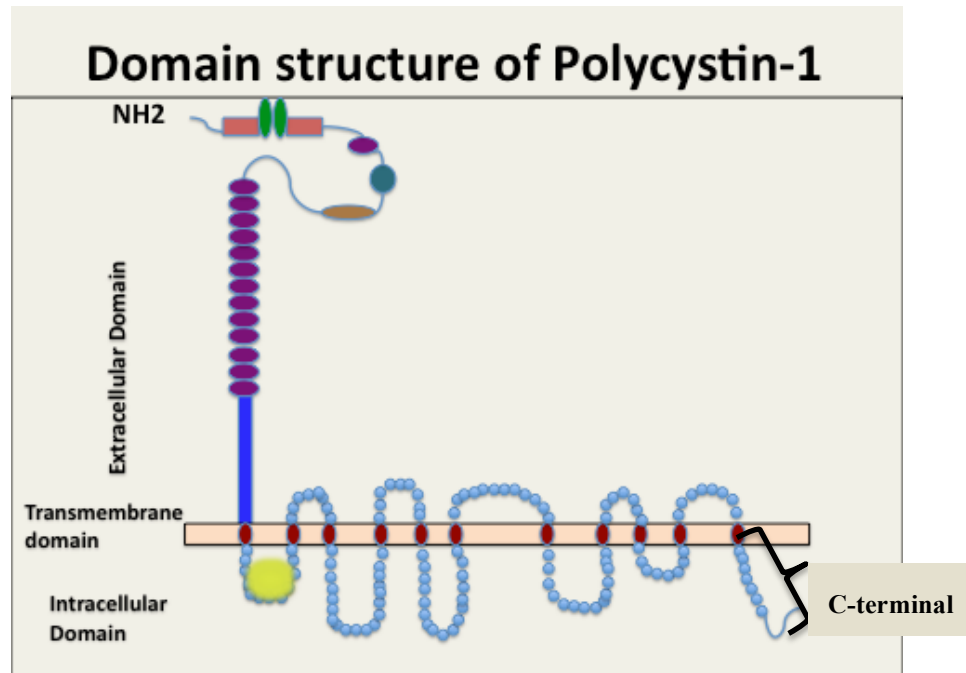


Figure 1.15: Domain structure of polycystin1. The Extracellular domain: cysteine-rich ■ leucine-rich repeats ● C-type lectin motif ● LDL-A ● Ig-like PKD ● REJ | Transmembrane region ● . Intracellular domain: PLAT domain ● .

The Polycystin 1 transmembrane domain consists of 11 transmembrane regions with intervening intracellular and extracellular loops (Hughes *et al*, 1995).

Intracellular Domain: The C-terminal of polycystin 1, (CT1) consists of approximately 200 amino acids. The sequence alignment of CT1 suggests homologs domains from different species that involve in protein-protein interaction and phosphorylation signaling (Wilson, 2004). The structure prediction of CT1 was obtained using Phyre2 (www.sbg.bio.ic.ac.uk/phyre2/), the predicted structure consists of three α helices with connecting loops (Figure 1.16).

1.4.2.2 Cellular and subcellular localization of Polycystin 1

Several studies have used anti-polycystin 1 antibodies and cDNA to investigate the tissue, cellular and subcellular distribution of polycystin 1. Polycystin 1 is highly expressed in various tissues including brain, liver, pancreas, heart and intestine, and it is also highly produced in the renal tubular epithelial cell of the kidneys (Adelsberg *et al*, 1997; Geng *et al*, 1996; Ward C *et al*, 1996). The cellular and subcellular patterns of polycystin 1 expression in the human and mouse kidney suggest that polycystin 1 is fairly concentrated in the epithelia of the ureteric bud (develops to adult kidney) during embryogenesis (Wilson, 2001).

1.4.2.3 Function of Polycystin 1

Although the polycystin 1 function in the kidney is not fully understood, there is strong evidence of a developmental regulation role for it in the kidney. This has been shown by mutations of the PKD1 gene in mice leading to cystic kidneys and embryonic or perinatal death (Lu *et al*, 1997) It has been deduced from a wide range of studies that polycystin 1 functions as a membrane receptor, capable of binding and acting with proteins by forming multiprotein complexes at focal adhesion, cell-cell junctions and cilia. Also it is able to bind and interact with carbohydrates and lipids. These interactions with the extracellular matrix or with other cell lead to intracellular signaling via a phosphorylation cascade and transcriptional regulation of proteins that regulate renal morphogenesis and differentiation (Wilson, 2001; Wilson, 2004).

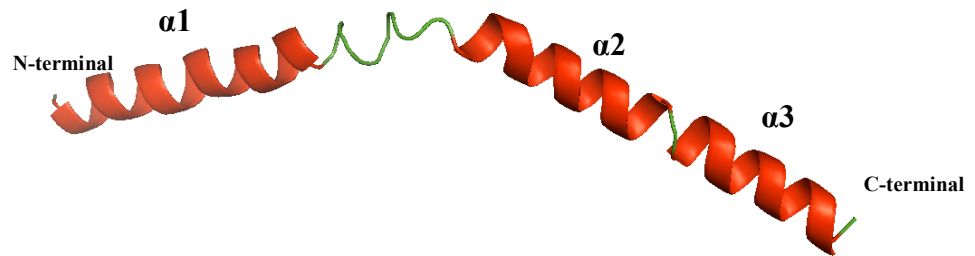


Figure 1.16: The structure prediction of the C-terminal domain of Polycystin1 (26 kDa). The structure consists of three α helices with connecting loops. The structure obtained by using Phyre2 (www.sbg.bio.ic.ac.uk/phyre2/).

1.4.3 Polycystin 2

1.4.3.1 Domain structure of Polycystin 2

The PKD2-encoded protein polycystin 2 is a 110 KD membrane protein. Sequential analysis using UniProt Predict shows that polycystin 2 consists of six transmembrane domains and that both N-terminal and C-terminal of polycystin 2 are intracellular domains (Figure 1.17). Polycystin 2 belongs to the Transient Receptor Potential (TRP) channel superfamily which is distinguished by six transmembrane domains and an ion pore between the fifth and sixth transmembrane domains (Luo *et al*, 2003).

Transmembrane Domain:

The Polycystin 2 transmembrane domain (Figure 1.17) contains six helices with connecting loops homologous to the last six transmembrane helices of polycystin1 (Mochizuki *et al*, 1996) As it shares homology with the Transient receptor potential TRP superfamily the cation channel is located between the fifth and sixth transmembrane helices.

The C-terminal domain of polycystin 2 CT2:

The C-terminal portion of polycystin 2, CT2, consists of approximately 289 amino acids. Sequence analysis suggests three interesting regions in CT2: an EF-hand domain in the N-terminal of CT2 (residues 720 to 796), a linker region (797-825) and an oligomeric coiled coil region (833-893) (Mochizuki *et al*, 1996). Structures of two domains from CT2 have been solved recently: the CT2EF-hand by NMR (Petri *et al*, 2010), and the CT2 coiled coil by X-ray crystallography (Yu *et al*, 2009). The study by Petri *et al* has described the NMR structure and the dynamics of Ca²⁺-bound CT2-EF. They found that human CT2-EF contains a divergent non-Ca²⁺-binding helix-loop-helix (α 1-loop- α 2) motif packed against a canonical Ca²⁺-binding helix-loop-helix (α 3-Ca²⁺-binding loop- α 4) motif (Figure 1.18) (Petri *et al*, 2010). The X-ray study of Yua *et al* (2009) on the C-terminal region of polycystin 2 solved the structure of a fragment from amino acids G833 to G895 at 1.9 Å resolution. It shows that this fragment forms a continuous α helix and assembles into a trimer (Figure 1.19). Also the study found that the C-terminal

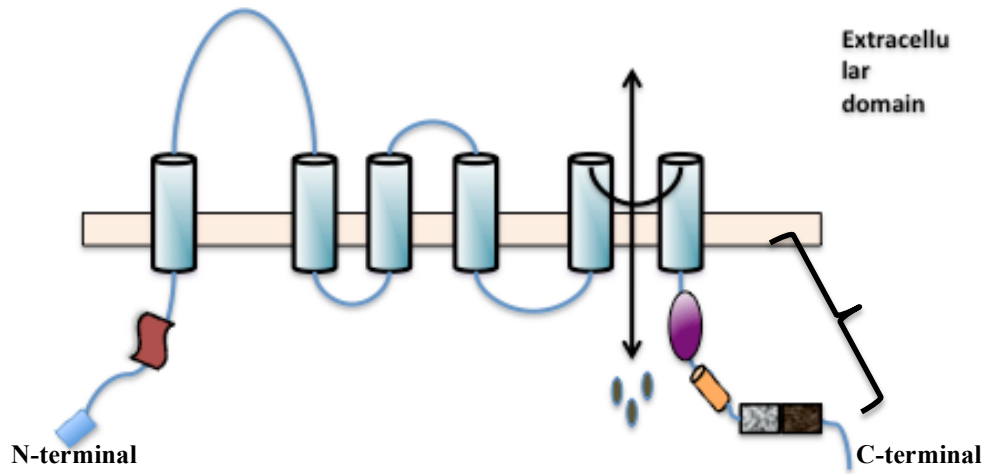


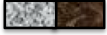


Figure 1.17: Domain structure of Polycystin 2. Six transmembrane helices and an ion pore between the fifth and sixth transmembrane domains. The intracellular domain contains the N- and the C-terminus. CT2EF-hand domain , Linker domain , CT2 coiled coil domain or oligomerization domain 

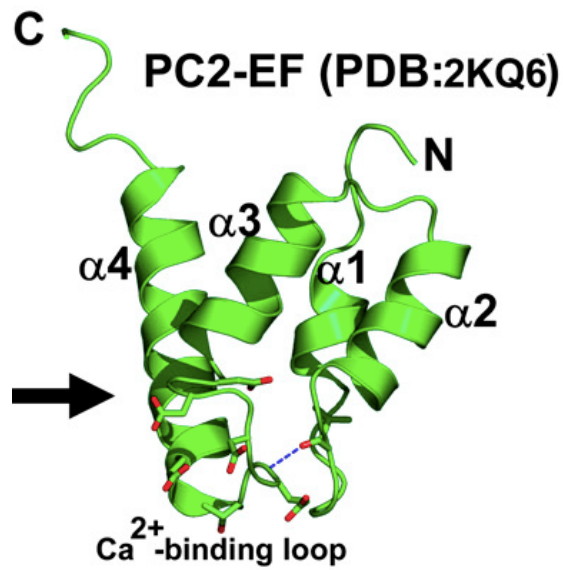


Figure 1.18: Human CT2-EF domain: This domain consists of two HLH (alpha helix-loop-alpha helix) motifs. The (α 1-loop- α 2) on the N-terminal is a non- Ca^{2+} -binding loop, and the (α 3-loop- α 4) at the C-terminal is a Ca^{2+} -binding loop. Adopted from (Petri *et al*, 2010).

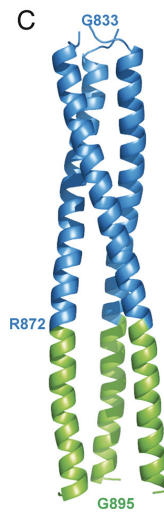


Figure 1.19: Human CT2 coiled coil domain: The CT2 coiled coil domain forms a trimer and also the C-terminal portion of α helices (A873 to G895) splays open to form an interaction region with another trimer (colored green). Adopted from (Yu *et al*, 2009).

portion of the α helices from A873 to G895, splay open and in the crystal lattice this region interacts with the same region of another trimer, forming a hexameric assembly (Yu *et al*, 2009). The structure prediction of CT2 was obtained using Phyre2 (www.sbg.bio.ic.ac.uk/phyre2/), the predicted structure consists of the EF-hand domain and two helices may be representing the coiled coil domain, and they connected by link loop (Figure 1.20).

1.4.3.2 Cellular and subcellular localization of Polycystin 2

Polycystin 2 is widespread in many tissues and it regulates the development and differentiation of the kidneys with the highest level in the renal tubular epithelium (Ong *et al*, 1999). Although there are some overlapping expression areas of polycystin 1 and polycystin 2, which suggests the possibility of interaction between them, there are some tissues in which polycystin 2 is expressed but polycystin 1 is not, indicating independent functions (Foggensteiner *et al*, 2000). Also at the subcellular level some similar tubular cell types are able to produce both proteins but their locations do not overlap with each other (Foggensteiner *et al*, 2000).

1.4.3.3 Function of Polycystin 2

In comparison with polycystin 1, polycystin 2 function is less well defined and studied. However some current results suggest that polycystin 2 can form heterodimers with polycystin 1 (Qian *et al*, 1997). The sequence homology between polycystin 2 and the transient receptor potential (TRP) has led to the prediction that polycystin 2 functions as a Ca^{2+} permeable nonselective cation channel (Luo *et al*, 2003).

1.4.4 Polycystin 1 and polycystin 2 interaction

Many studies have suggested that the presence of coiled coil motifs in the C-terminal tails of both polycystin 1 and polycystin 2 may allow protein-protein cytoplasmic interaction (Wilson, 2001). The coiled-coil domain of polycystin 1

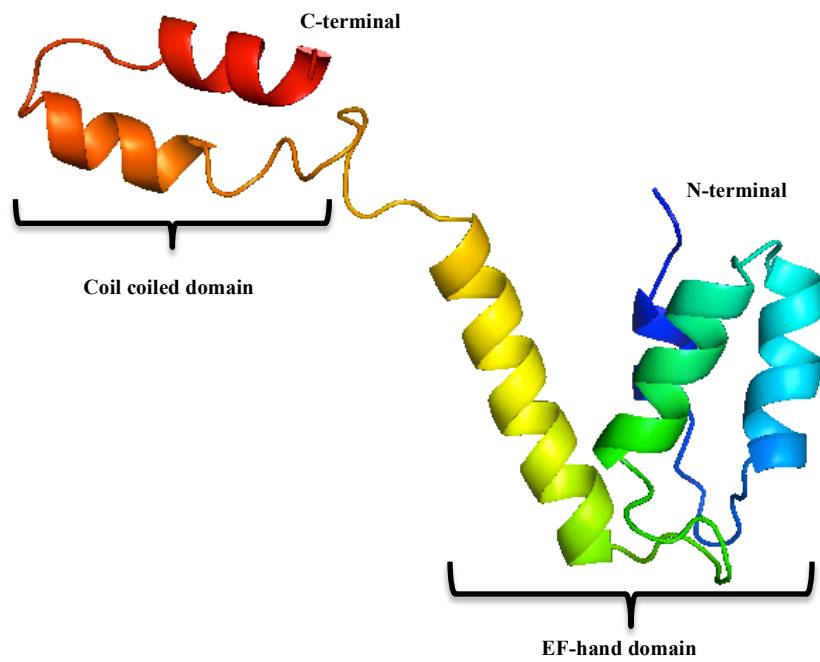


Figure 1.20: The structure prediction of the C-terminal domain of Polycystin 2 (35 kDa). CT2 is composed of the EF-hand domain on the right side and the coil-coiled domain on the left. The 3D structure was predicted from the protein amino acid sequence by using Phyre2 server <http://www.sbg.bio.ic.ac.uk/phyre2/>.

has been shown to bind polycystin 2 *in vitro* in experimental analyses using yeast two-hybrid and overexpression techniques (TSIOKAS *et al*, 1997). Immunolocalization studies have been used to investigate the polycystin 1-polycystin 2 interaction *in vivo*, suggesting that polycystin 1 and polycystin 2 occasionally co-localize but they often do not (Foggensteiner *et al*, 2000).

1.5 Aims of the project

Given that the autosomal dominant polycystic kidney disease is the most common human genetic disease, with a prevalence of 1 in 800 of the world's population, the Polycystin 1 and Polycystin 2 proteins became important targets for many laboratories. The determination of the 3D structures of these proteins may be a significant target for researchers to increase the understanding of their functions. However, from the point of view of the *E. coli* protein overexpression system, which is the easiest, quickest, and most commonly used procedure, the overexpression of a large transmembrane protein such as Polycystin 1 or Polycystin 2 would be impractical. Thus many structural studies focus on the domains that made up the full protein, and together these studies can create an image of the whole protein. It has been thought that the C-terminal domains of Polycystin 1 and Polycystin 2 play a role in protein-protein interactions but the 3D structures of these domains remain undefined. In order to investigate the roles of the C-terminal domains of Polycystin 1 (CT1) and Polycystin 2 (CT2), this project aims to obtain the 3D crystal structure of each C-terminal domain. The primary aim of this project was to overexpress, purify, crystallize and determine the 3D structures of CT1 and CT2. The interaction between CT1 and CT2 has shown *in vitro*, several studies suggest that the interaction between Polycystin 1 and Polycystin 2 may be regulated through the coiled-coil domains (Wilson, 2001). Therefore a secondary goal of this project was to achieve the 3D crystal structure of CT1 in complex with CT2.

Chapter 2: Materials and Methods

This chapter includes a general description of the methods and techniques used during the work in these two projects. Also this chapter describes the reagents and equipments used in this thesis. The work in all two projects in general is aimed at the crystallization of these proteins and obtaining the 3D structures using X-ray crystallography. X-ray structural studies demand large quantities of soluble, properly folded pure protein. But it has become clear that obtaining these proteins from their natural sources is often impractical, due to low of expression yields. Therefore use of recombinant DNA technologies is an alternative feasible application that is able to produce large amounts of a desired protein. The technique involves extraction of the DNA from the organism being studied, followed by amplification of the gene of interest by use of PCR. The gene is then inserted into an expression plasmid and transformed into a host that can be manipulated into producing large amounts of the protein of interest.

2.1 Bacterial strains, DNA, plasmids and oligonucleotides

- **Bacterial strains:** The following *E. coli* strains were all purchased from Novagen; *E. coli* BL21(DE3), *E. coli* BL21(DE3)-RIPL, *E. coli* Novablue and *E. coli* Tuner DE3 pLacI.

- **Genomic DNA:** *Burkholderia pseudomallei* strain D286 genomic DNA, taken from a melioidosis patient at Kuala Lumpur Hospital.

- **Plasmids:** The following plasmids were used in this thesis:

1. The recombinant C-terminal polycystin2 plasmid His-CT2 pET-28a(+)

Dr Yaoxian Xu from Prof Albert Ong's team (Academic Unit of Nephrology, Department of Infection and Immunity The University of Sheffield Medical School) has supplied this plasmid.

4. The recombinant C-terminal polycystin1 plasmid GB1-CT1 GEV-S2

Also provided by prof Albert Ong team.

5. pETBlue-1 vector

The vector was purchased from Novagen

- **Oligonucleotides:** The oligonucleotides used during the course of this thesis were synthesized by Eurofins MWG Opera.

1. BPSL2418 forward primer

BPSL2418-F: 5' ATGTTTCGCGCTTTCCGAAGC 3'

2. BPSL2418 reverse primer

BPSL2418-R: 5' GATGCCGATTTACTCGCCGC 3'

3. pETBlue-1 up: 5' TCACGACGTTGTAAAACGAC 3'

4. pETBlue-1 down: 5' GTTAAATTGCTAACGCAGTCA 3'

2.2 Materials

Most of the chemicals, reagents and apparatus used in this thesis were common materials and are extensively offered by a wide range of producers. Chemicals were purchased from BDH Laboratory Supplies, Poole, England; Fisher Scientific UK Ltd., Leicestershire, UK; Bio-Rad Laboratories, Inc., CA, USA; Sigma-Aldrich, St. Louis, MO, USA; Invitrogen Corporation, CA, USA; Qiagen, Hilden, Germany; and GE Healthcare, Freiburg, Germany.

Substrates: Free Met-SO (L-Met-*R*-sulfoxide) from Sigma-Aldrich, St. Louis, MO, USA

2.3 Microbiological Methods

2.3.1 Growth media

1. Luria-Bertani (LB)

Bacterial strains were grown in sterilized Luria-Bertani (LB) medium (Sambrook & Russell, 2006).

The table below describes the composition of media used for *E. coli* culture.

	LB medium	LB-agar
Tryptone	10 g/l	10 g/l
Yeast extract	5 g/l	5 g/l
NaCl	10 g/l	10 g/l
Bacteriological agar	-	15 g/l

2. The SOC medium

This growth medium was used in the transformation of competent cells with recombinant plasmid in the BPSL2418 project (Sambrook & Russell, 2006).

The table below describes the composition of media used for *E. coli* culture.

	Amount for 1L
Tryptone	20 g
Yeast extract	5 g
NaCl	0.5 g
1M KCl	2.5 ml
1M MgCl ₂	10 ml
1 M MgSO ₄	10 ml
ddH ₂ O	To 1000

All media were sterilized by autoclaving, and desired antibiotics were added.

The table below describes the antibiotics used and their working concentrations

Antibiotic	Final antibiotic concentration
Ampicillin (Amp)	100 µg/ml
Chloramphenicol (Cam)	75 µg/ml
Kanamycin (Kan)	50 µg/ml
Carbenicillin (Car)	50 µg/ml
Tetracycline (Tet)	15 µg/ml

2.4 Polymerase chain reaction (PCR)

PCR (Bartlett & Stirling, 2003) was used for two different applications during this work: (1) amplification of inserts for cloning and (2) colony PCR to confirm the insert's presence in the purified plasmid post-cloning. In both applications the following reaction mixture was used. Occasionally the PCR mixture was varied by the addition of 10% DMSO to reduce primer secondary structure formation:

Genomic template (~ 2 µM)	1 µl
Sense primer (10 pM)	1 µl
Antisense primer (10 pM)	1 µl
BioMix	25 µl
Sterile, deionized water	22 µl
Total volume	50 µl

BioMix is a premixed PCR cocktail containing *Taq* DNA polymerase, reaction buffer and free dNTPs available from Bionline.

The PCR reaction mixture was set in thermal cycling using the protocol below:

Initial denaturation	94 °C	4 mins	} 30 cycles
Denaturation	94 °C	1 min	
Annealing	55 °C	1 min	
Extension	72 °C	1 min	
Final extension	72 °C	10 mins	
Final hold	4 °C		

In order to carry out colony PCR the genomic template was changed by boiled cell-lysate from colonies picked from an agar plate resuspended in 30 µl sterile water.

2.5 Agarose gel electrophoresis

Agarose gel electrophoresis was used for the separation of DNA fragments to either analyze the PCR products or to confirm the presence of inserted genes in recombinant constructs. This procedure was accomplished by running samples on 1% agarose gels and visualizing their migration through the gel matrix. Gel formation was carried out by melting 0.5 g agarose into 50 ml of TAE buffer (40 mM Tris pH 8.0, 20 mM glacial acetic acid, 1 mM EDTA) into a 200 ml Duran using a microwave. When the melted agarose cooled down enough to be easily touched by hand, a 5 µl of x10,000 GelRed or ethidium bromide was added to permit fluorescent visualization of DNA migration under UV light. The gel was poured into BioRad Mini-Sub Cell gel-bed (gel rack) with a comb stuck in to form the wells. The gel was allowed to set for approximately 20 minutes. Once it had solidified the gel was submerged in TAE buffer and the samples, combined with loading buffer (0.25% bromophenol blue, 30% glycerol) were loaded into the formed well. In order to determine the molecular weight of DNA fragments, a molecular weight marker (Hyperladder I from Bionline) was loaded. Electrophoretic fractionation was carried out by connecting the gel-tank to its correctly oriented electrodes and was run at 100 V. Times ranged between 40-70 minutes depending on DNA size, where smaller PCR products were run for 40 minutes and larger

fragments such as digested plasmids were run for 60-70 minutes. The gel was visualized using a UV lamp and photographs were taken using a digital camera.

2.6 Purification of PCR products by gel extraction:

The following protocol was used to extract and purify the DNA of PCR products from an agarose gel using a QIAquick Gel Extraction kit and a microcentrifuge:

The DNA fragments were cut from the agarose gel with a clean sharp scalpel and placed in 1.5 ml tubes. Gel slices were weighed, and for gel solubilisation and DNA binding to the column's membrane, 3 volumes of buffer QG (from Qiagen Gel Extraction Kit which contains Guanidium Thiocchloride), were added to 1 volume of gel. Gel tubes were incubated at 50 °C for 10 minutes or until the gel slices had totally dissolved. The mixture was placed in a QIAquick spin column. To bind DNA, 1 gel volume of isopropanol was added and centrifuged for 1 minute at 13000 rpm. 0.5 ml of QG buffer was added and centrifuged for 1 minute. 0.75 ml of ethanol-containing PE washing buffer was added and centrifuged for 1 minute. The flow-through was discarded and an additional 1 minute of centrifugation was run to remove the remaining buffer. The QIAquick spin column was replaced into a clean 1.5 ml tube. To elute the DNA, 30 µl of EB buffer (10 mM Tris-HCl, pH 8.5) was added to the center of QIAquick membrane, the column was allowed to stand for 2-5 minutes, and then centrifuged for 1 minute.

2.7 Plasmid purification

All plasmids used in this thesis were extracted from their *E. coli* hosts by applying the standard protocol of Plasmid DNA Purification using the QIAprep Spin Miniprep Kit from (QIAGEN) and a microcentrifuge. A colony of *E. coli* cells, containing a particular plasmid vector growing on LB agar containing 100 µg/ml carbenicillin, was picked into 3 ml of LB media containing the appropriate antibiotics and grown overnight at 37 °C at 200 rpm. Cells were then harvested by centrifugation at 5,000 g for 20 minutes before the plasmids were extracted. Cell pellets were resuspended in 250 µl buffer P1, this buffer contains RNase to degrade

any RNA, preventing it contaminating the purified plasmid. The cells were then subjected to alkaline lysis by the addition of 250 μ l buffer P2 and mixing by gentle inversion. The solution was neutralised and the salt concentration adjusted to allow binding to a QIAprep spin column by the addition of 350 μ l buffer N3. DNA adsorbs to the silica membrane of the spin column in the presence of high concentrations of chaotropic salts at a pH below 7.5. The lysate is then cleared of cell debris by centrifugation at 17,000 g for 10 minutes in a benchtop centrifuge. All subsequent centrifugation was done at 17,000 g for 60 seconds. The supernatant was applied to a QIAprep spin column, centrifuged and the flow-through discarded. A wash was carried out using 750 μ l buffer PE, which contains ethanol to precipitate the DNA on the column while removing the salts. The column was centrifuged, the flow-through discarded and the column was again centrifuged to ensure all ethanol had been removed from the column. The column was then transferred to a clean microcentrifuge tube and the DNA was eluted by adding 65 μ l of water, incubating at room temperature for 2 minutes and finally centrifuging to collect the pure plasmid DNA. Yields from minipreps typically vary between 10 and 100 ng/ μ l.

2.8 Cloning of expression construct

The BPSL2418 expression construct (*bpsl2418* pETBlue-1) was cloned during this thesis. The general methods and materials used in cloning this construct will be described below, but more details are given in chapter 4.

2.8.1 Cloning with pETBlue-1 vector

The plasmid pETBlue-1 vector (Figure 2.1) facilitates the expression of native unfused proteins and allows convenient subcloning of target genes already fused to existing detection and purification tags. The EcoRV cloning site is appropriately spaced down stream of an *E. coli* ribosome-binding site. The insert must encode an ATG start codon at its 5' end if expression is desired. The pETBlue-1 vector allows insertion of the target gene through blunt ended cloning into an EcoRV restriction site present in a copy of the α -peptide fragment of the *lacZ* gene. The

pETBlue-1 sequence landmarks	
lac operator	3429-3448
T7 promoter	1-17
lac operator	22-42
T7 transcription start	18
multiple cloning region (EcoR V-SrfI)	276-297
lacZ start codon	314
lacZ α -peptide ORF	57-314
E. coli promoter	364-392
f1 origin	919-1374
bla coding sequence	1492-2349
pUC origin	3029

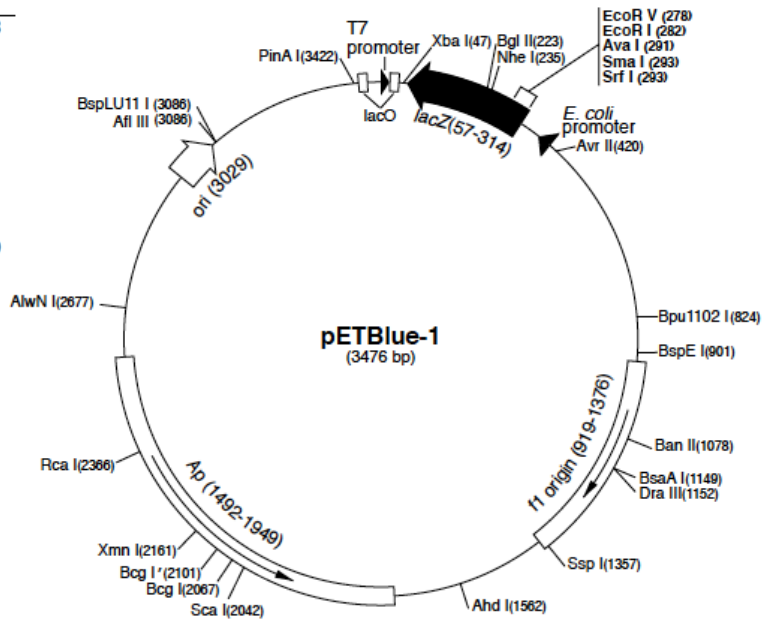


Figure 2.1: pETBlue1plasmid map.

Adapted from Novagen (<http://www.merckmillipore.co.uk>)

produced plasmid can then be transformed into a bacterial strain deficient in the T7 polymerase gene to restrict any background expression and with a copy of the ω -peptide gene of *LacZ* in its genome. Blue/white screening on agar containing IPTG and X-gal can be used to select colonies containing a plasmid with an insertion. The *lacZ* gene fragments are responsible for β -galactosidase expression, which allows the break down of X-gal, generating a blue pigment. If the *LacZ* α -peptide gene does not contain an insertion, it is expressed and can combine with the ω -peptide of *LacZ*, produced by expression from the genome, giving a functional copy of β -galactosidase and a blue colony. Plasmids with an insertion give white colonies because they are able to prevent the expression of the α -peptide and therefore β -galactosidase activity. In protein expression, expression strains that contain a copy of *lacI* gene such as the pLacI strains should be used in combination with the plasmid (pETBlue-1), because it is lacking a copy of this gene. The pLacI strains can be selected by a chloramphenicol resistance marker.

2.8.2 Ligation and transformation for pETBlue-1 cloning

Ligation reactions were set up containing 2 μ l purified PCR product (approximately 50 ng) and 50 ng AccepTor vector in Clonable™ ligation premix buffer (Novagen) and were incubated at 16 °C for 30 minutes. Ligation reactions were then used to transform Novablue *E. coli* cells. Eppendorf tubes containing 50 μ l aliquots of cells were removed from the -80 °C freezer and incubated on ice for 5 minutes. Once defrosted 1 μ l of the ligation reaction was added before being left to incubate on ice for 5 minutes. The cells were subjected to a heat shock at 42 °C for 30 seconds before being returned to ice for 2 minutes. 250 μ l of SOC media was then added and the cells were then incubated at 37 °C at 200 rpm for 60 minutes before plating on LB agar containing 100 μ g/ml carbenicillin, 15 μ g/ml tetracycline, 70 μ g/ml X-gal and 80 μ M IPTG for selection and blue-white screening of colonies to select for transformants, that had white colonies containing an insert in the vector.

2.8.3 Confirmation of cloning results

White colonies were selected and transferred into 100 µl water and boiled at 100 °C for 10 minutes. The mixture was centrifuged at 17,000 g for 5 minutes to remove cell debris. PCR conditions were set up as mentioned earlier in (2.4). Results from pETBlue-1 cloning were confirmed using two reactions for each colony using pETBlueDOWN or pETBlueUP. PCR products were analyzed by electrophoresis on a 1% TAE agarose. Desired colonies with correctly sized band(s) were selected and transferred into 3 ml LB including 100 µg/ml ampicillin and grown overnight at 37 °C at 200 rpm. The cells were harvested and plasmids were purified using a QIAprep Spin MiniPrep Kit (QIAGEN) and the same protocol as before in (2.7). Purified plasmids were sent for sequencing using T7F and T7R primers (SourceBioscience or Geneservice).

2.9 Transformation

Recombinant plasmids were transformed into overexpression hosts such as *E. coli* BL21(DE3), *E. coli* BL21(DE3)-RIPL and *E. coli* Tuner DE3 pLacI (Tu *et al*, 2005). Eppendorf tubes containing 20 µl aliquots of cells at -80 °C were defrosted by incubating the tubes on ice for 5 minutes. 1 µl of plasmid DNA (approximately 10 ng/µl) was added to the defrosted cell and the reaction was incubated on ice for 5-30 minutes. The cells were exposed to a heat shock at 42 °C for 30-60 seconds before incubating again on ice for 2 minutes. 250 µl of SOC media was then added in case of transformed plasmid into Tuner DE3 pLacI while LB media was added in case of *E. coli* BL21(DE3) and *E. coli* BL21(DE3)-RIPL. The cells were incubated at 37 °C at 200 rpm for 30 minutes but the Tuner DE3 pLacI was incubated at 37 °C at 250 rpm for 60 minutes before plating on LB agar containing the desired antibiotics to select for successful transformants.

2.10 Protein expression

The *E. coli* expression system was used to express BPSL2418, CT1 and CT2. Optimization work was performed to produce soluble protein. Most of the proteins in this thesis were successfully produced as a large amount of soluble proteins but

in a single case the protein expressed as inclusion bodies. Several attempts at protein refolding were made in order to achieving soluble proteins.

2.10.1 Regulation of protein expression in the pET expression system

The expression hosts used in this thesis are all DE3 lysogen strains containing a copy of T7 prophage DNA that includes the T7 polymerase gene downstream to a *lac* promoter site and *lac* operator-binding site. These strains also have a copy of the *lacI* gene that produces *lac* repressor protein. The lack of lactose (under normal conditions) permits the binding of the *lac* repressor to the *lac* operator sequence restricting the T7 polymerase expression. Inducing the culture by IPTG (a non-hydrolysable lactose analog), the IPTG binds to the *lac* repressor preventing it from binding to the operator sequence and permitting the T7 polymerase gene transcription. Desired genes are cloned into pET vectors downstream of a T7 promoter and *lac* operator site. When T7 polymerase is produced, it transcribes a gene within a pET vector resulting in expression of the target gene. The plasmids pET-28a(+) and GEV-S2 (vector backbone pET21a) have an extra copy of the *lac* repressor gene to prevent basal expression without the induction of IPTG. The absence of the *lac* repressor gene in pETBlue-1 imposes the use of strains, which possess an additional copy of the gene encoded on a plasmid, such as the pLacI strains.

2.10.2 Protein overexpression protocol

In general a similar procedure has been used to express all proteins involved in this thesis, where changes have been done to optimize expression yield, there will be mentioned in the following chapters.

2.10.2.1 Small-scale overexpression trials

An initial overnight 5 ml culture grown in LB containing suitable selectable antibiotics was used to inoculate 500 ml LB (1% culture dilution) containing the

same antibiotics in primary culture. In some cases 1% glucose was added to LB culture to increase growth. The culture was grown at 37 °C and 200-250 rpm until the optical density at 600 nm reached 0.5-0.8 at which point the culture was divided into 50 ml aliquots in 250 ml flasks. Each individual culture was induced by the addition of IPTG and treated with various post-induction conditions. Different conditions were prepared by changing in temperature, induction time and IPTG concentration in order to produce a soluble overexpressed protein, e.g. temperature (18 – 37 °C), time (3 – 20 hours) and IPTG concentration (0.3 – 1 mM). Samples were taken before and after 1, 2, 3, 4, 18 and 20 hours of induction to determine the optimum condition for each protein. Samples were centrifuged at 17,000 g for 5 minutes in a bench top centrifuge. Samples then stored at -20 °C before cell lysis using bugbuster (Novagen) and PAGE analysis. A mixture of 100 µl bugbuster A with 1 µl bugbuster B (Novagen) was used to resuspend the cell pellets and then the resuspended cell were incubated at room temperature for 15 minutes. Cell debris and insoluble protein were discarded by centrifugation at 17,000 g for 10 minutes and the supernatant (the soluble fraction) was collected. The pellet, which includes the inclusion bodies, was resuspended in 100 µl 4% SDS and incubated at room temperature for 15 minutes. The resuspended pellet was centrifuged at 17,000 g for 10 minutes and the supernatant was collected as the insoluble fraction. The different overexpression conditions were then judged by running fractions of each condition on polyacrylamide gels to identify the optimum conditions.

2.10.2.2 Large-scale overexpression trials

Once optimum conditions were determined for each protein the overexpression culture was scaled up to between 2 and 4 l of media in 500 ml aliquots in 2 l conical flasks. Cells resulting from overexpression were harvested by centrifugation at 5,000 g for 45 minutes and then resuspended in approximately 70 ml LB. The cells were aliquoted in 50 ml falcon tubes and centrifuged again at 5,000 g for 30 minutes to remove any excess media. The supernatant was removed and the cell pellets were stored at -20 °C until required.

2.11 Sodium dodecyl sulfate Polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used routinely throughout this thesis for analysis of proteins and to evaluate the quality of protein expression and purity. SDS-PAGE was carried out according to the protocol mentioned in Laemmli's study (Laemmli, 1970). All PAGE equipment was purchased from BioRad and all gels were stained with coomassie Brilliant Blue. Gel and buffers ingredients used in this thesis were taken from 'Manual of Molecular Cloning' (Sambrook & Russell, 2006). All SDS-PAGE gels reported in this thesis were 12% (acrylamide concentration).

12% SDS-PAGE separating gel:

Acrylamide/ Bis solution	5 ml 30% (w/v)
1 M Tris-HCl, pH 8.8	4.69 ml
10% SDS (w/v)	125 μ l
Milli-Q water	2.56 ml
Ammonium persulphate	125 μ l 10% (w/v)
N,N,N',N'-tetramethyl-ethane-1,2-diamine (TEMED)	12.5 μ l

6% SDS-PAGE stacking gel:

Acrylamide/ Bis solution	1.5 ml 30% (w/v)
1 M Tris-HCl, pH 6.8	0.94 ml
10% SDS (w/v)	75 μ l
Milli-Q water	2.6 ml
Ammonium persulphate	75 μ l 10% (w/v)
N,N,N',N'-tetramethyl-ethane-1,2-diamine (TEMED)	7.5 μ l

1x SDS-running buffer:

Glycine	14.4g /l
Tris-HCl	3g /l
SDS	1g /l
*Adjust to pH 8.8	2.56 ml

2x SDS-loading buffer:

Glycerol	20% (v/v)
Tris-HCl	100 mM
SDS	4% (v/v)
Bromophenol blue	0.02% (v/v)
β -mercaptoethanol	200 mM

1x Coomassie blue stain:

Methanol	20% (v/v)
Acetic acid	7.5% (v/v)
Coomassie Blue	4% (v/v)
Bromophenol blue	0.1% (w/v)

Destain:

Methanol	20% (v/v)
Acetic acid	7.5% (v/v)

2.12 Solubilisation and refolding of proteins

As mentioned above, CT1 was overexpressed as insoluble inclusion bodies in *E. coli*. The inclusion bodies were isolated and attempts at protein solubilisation and refolding were performed.

2.12.1 Inclusion bodies preparation

The following method illustrates the treatment of CT1 inclusion bodies from 5 l of *E. coli* culture. The pelleted cultures were dissolved on ice and resuspended in lysis buffer (10 mM Tris, 1 mM EDTA, 10 mM DTT, pH 8.0 + protease inhibitor cocktail tablet in a 1:5 ^{w/v} ratio). In order to inhibit proteolysis a 1 mM PMSF protease inhibitor was added, and to break the bacterial cell walls 500 µg/ml lysozyme was added. This suspension then was kept on ice for 30 minutes, with shaking by inversion every few minutes. The suspension was divided into 3-4 Sorval centrifuge tubes (each tube contained ~10 ml). Cell disruption was carried out using the sonication method, where each tube of all suspension was sonicated on ice for 60 seconds, five times, with 4 minutes rest on ice between each 60 seconds sonication. Sonication was run at the maximum attainable amplitude. The suspension was centrifuged at 40,000 g for 30 minutes and the supernatant was removed. Each pellet was resuspended in 25 ml of ice-cold wash buffer (50 mM Tris, 1 mM EDTA, 10 mM DTT, 2% ^{w/v} sodium deoxycholate, pH 8.0). To enhance the pellet homogeneity, an additional 2 x 60 seconds of sonication was applied to the resuspended pellet (as mentioned above). The suspension was centrifuged at 40,000 g for 30 minutes and the supernatant was removed. The last step (washing and centrifugation) was rerun for four times, saving the pellet each time. To discard residual detergent, an additional washing was done using distilled water instead of wash buffer, a centrifugation step (40,000 g) for 60 minutes was needed to completely extract the inclusion bodies from buffer and detergent. A final wash step using distilled water and sonication step (30 sec) were accomplished to sufficiently disperse the pellet. Inclusion bodies were saved at 4 °C for solubilisation the next day or at -80 °C for long-term storage.

2.12.2 Solubilisation and refolding of inclusion bodies

CT1 was refolded using a general protocol mentioned in the literature (Sandowski *et al*, 2002).

2.12.2.1 Refolding CT1

This protocol explains the refolding of CT1 inclusion bodies produced from 2.5 l of *E. coli* culture. The CT1 inclusion bodies were denatured using a solubilisation buffer (100 mM Tris, 8 M urea, 500 mM L-arginine, 10 mM cysteine, pH 8.0), which was prepared on the day of use. 2 g of TMB mixed bed resin (Sigma-Aldrich) was added to deionise the solubilisation buffer, this step was performed at 4 °C in the dark for 2 hours. The buffer was filtered before the solubilisation step. CT1 inclusion bodies were added to 500 ml of ice-cold solubilisation buffer, and the pH adjusted to 11.3, the mixture was stirred at 4 °C in the dark for 2 hour. The mixture was transferred to dialysis tubing and dialysed against 10 l of ice-coled refolding buffer (10 mM Tris, pH 8.0 at 4 °C). CT1 refolding was accomplished for a course of 48 hours with 6 external buffer changes (4 hours between each change). After 48 hours the solution was centrifuged (13000 rpm for 10 mins) to pellet the refolded protein and to discard the buffers. The protein pellet was diluted with 50 mM Tris pH 8.8. The refolded protein was immediately used for purification.

2.13 Protein purification

Protein crystallization procedures demand pure, homogeneous and concentrated (regularly between 5 and 30 mg/ml) protein. In order to achieve pure concentrated proteins, several types of column chromatography were used throughout this thesis. The first step in the purification of all desired proteins was the lysing of the cells that contain the target protein. This was achieved by the procedure of sonication, which utilizes high frequency of sonic pulses to destroy the membranes and cell walls of bacteria. Cell pellets frozen at -20 °C were defrosted and resuspended in a suitable buffer before homogenization by sonication. Sonication was run on ice for

20 seconds 3 times with 20 seconds rest on ice between each sonication, using the maximum attainable amplitude. Insoluble proteins and cell debris was separated by centrifuging the crude cell extract at 70,000 g for 15 minutes or 60,000 g for 20 minutes before purification. The following purification methods were used in this thesis:

2.13.1 Ion exchange chromatography

Proteins are separated using ion exchange chromatography on the basis of their overall net surface charge. The net surface charge of a protein is determined by the amino acids located on its surface and it also dependent on its pH (Hames & Hooper, 2005). If the isoelectric point (pI) of a protein is equal to the pH of its environment there is no overall charge on its surface. The protein has positive net surface charge if the pH is below the pI and it has negative charge if the pH above. According to this, ion exchange chromatography can be divided into two types: in cation exchange chromatography, the column includes negatively charged beads to bind a protein that has a positive charge. The second type is anion exchange chromatography, where the column contains positively charged beads to bind a negatively charged protein. Protein with a net surface charge opposite to the charged column bind, and then these proteins can be eluted by washing the column with increasing concentrations of a solution of sodium chloride. Na⁺ and Cl⁻ ions compete for the charged groups on the column leading proteins to elute from the column. In this thesis, columns including positively charged diethylaminoethyl (DEAE) were used such as DEAE Hi Trap 108 (anion exchange column DEAE-sepharose Fast Flow weak anion exchanger) from GE Helthcare and Resource Q (HPLC) (strong anion exchanger) from GE Healthcare.

2.13.2 Hydrophobic interaction chromatography

Proteins separate in this type of purification according to diversity in their hydrophobic groups on their surface such as phenyl, hexyl, butyl or ethyl groups. The existence of an anti-chaotropic salt such as ammonium sulphate increases the interaction between the hydrophobic groups on the protein surface and the

hydrophobic groups linked to the matrix by competing for dissolution. Protein solutions are loaded onto a hydrophobic column in a high concentration buffer, permitting hydrophobic patches on the surface of a protein to interact with hydrophobic groups of the beads leading to the binding of the protein to the column. To elute proteins from the column, the concentration of the buffer is reduced, the protein with highest hydrophobicity eluting last. In this work, the hydrophobic column used was a 5 ml Hi Trap Phenyl HP cartridge from GE Healthcare. A reverse gradient of ammonium sulphate concentration was used to elute proteins.

2.13.3 Gel filtration

Gel filtration is a technique that separates a mixture of proteins based on their size and shape by passing them through a column filled with porous beads. The volume inside the column is in two parts, the excluded volume outside the beads, and included volume inside the beads. Molecules larger than the pores are unable to enter the beads and can therefore only occupy the excluded volume whereas molecules that are small enough to enter the beads are able to occupy both the included and excluded volumes. Therefore larger molecules will elute first from the column, and as the size of a molecule decreases it can enter a larger proportion of the beads, retarding its progress down the column and causing it to elute later, until finally proteins that can enter all beads elute last. Assuming a protein is globular there is a linear relationship between the elution volume of a particular molecule from a given column and the logarithmic value of its molecular weight based on its partition coefficient:

$$K_{av} = \frac{\text{Eluted volume} - \text{Void volume}}{\text{Total Volume} - \text{Void volume}}$$

Therefore gel filtration can also be used to estimate the molecular weight of a protein by comparing its partition coefficient, K_{av} , to a calibration curve of partition coefficient plotted against the log of molecular weight for a particular gel filtration column determining its oligomeric state. Gel filtration was used in this thesis as

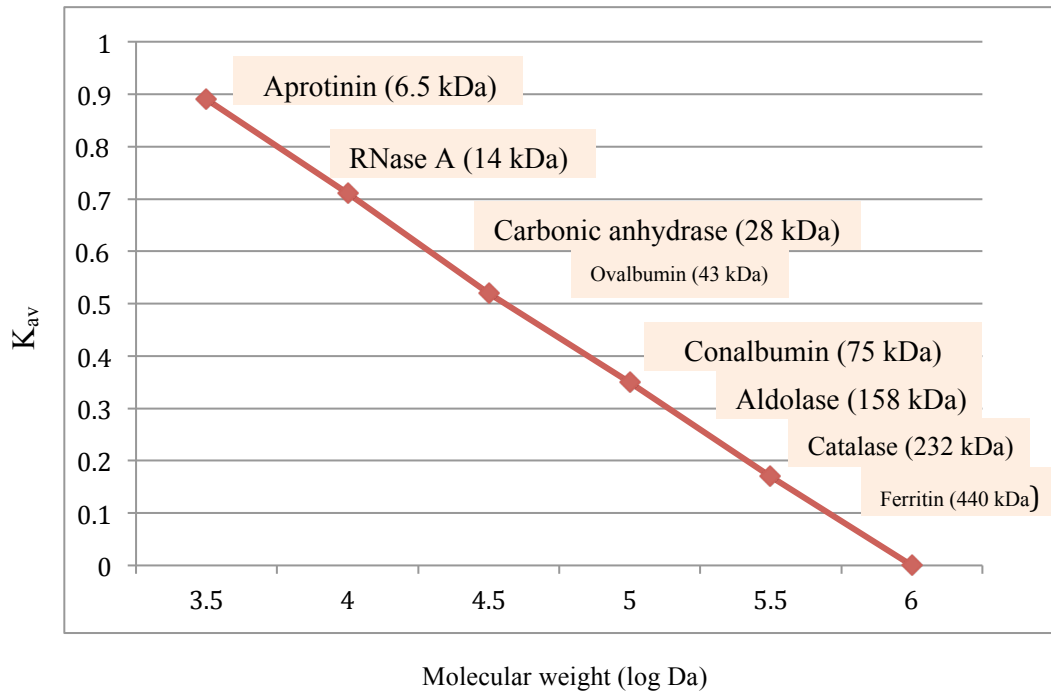


Figure 2.2: Calibration curve for gel filtration superdex 200 column.

purification step for all proteins, it was also used to determine the oligomeric state of CT2 (details given in chapter 7). A Hi-Load Superdex 200 1.6x60 cm column from GE Healthcare, was the gel filtration column in this work.

2.13.4 Nickel Nitrilotriacetic acid (Ni-NTA) Sepharose

The Ni-NTA is an affinity chromatography method for purifying recombinant, histidine tagged proteins. In this type of column, histidine residues in a tagged protein bind to the immobilized nickel ions in sepharose porous bead of the column. The histidine tag binds to the column with high specificity and affinity, and the non his-tagged proteins pass through the sepharose matrix. The bound protein is eluted from the column by adding increasing gradient of imidazole, which has a higher affinity for the metal more than the histidine affinity to the metal.

2.14 Protein concentration

Protein concentration is a critical factor for crystallization. Concentration was achieved using centrifugal ultrafiltration concentrators (Vivascience). This device contains two chambers and a polyethersulfone membrane between them. The membrane has pores of controlled size; proteins smaller than the pore size will pass from the first chamber to the second one. This device was available in different sizes such as 10, 30 or 100 kDa. Protein concentration was determined using a BioRad protein assay protocol of the Bradford method (Bradford, 1976). Reagents used in this protocol were purchased from BioRad. In this protocol, protein molecules bind Coomassie Brilliant Blue G-250 dye causing a change in the solution colour. The changing in colour was measured by using a spectrophotometer. The absorbance maximum for the dye changes from 465 nm to 595 nm when binding to a protein. Protein concentration was determined from the absorbance at 595 nm.

2.15 Protein crystallization to structure identification

The procedures for producing crystals, crystal mounting, data collection and solving a protein structure will be explained in chapter 3. This section describes the laboratory techniques employed to produce crystals and solve the structures for the target proteins.

2.15.1 Producing crystals

Initial crystallization screening trials for proteins were set up using a Matrix Hydra II Plus One crystallization robot (Figure 2.3). In event of the robot not functioning, the initial crystallization screens were set manually. Screens were performed in 96-well plates where each plate consumes a total of 20 μ l of protein, with 200 nl for each condition well. Different crystallization screens were used (JCSG suite, PACT suite, Classics suite, PEG suite, $(\text{NH}_3)\text{SO}_4$ suite and pH clear suite) from QIAGEN and (Crystal screen 1, crystal screen 2) from Jena Bioscience. Conditions producing promising crystals were optimized to generate desired size and quality of crystals. Optimization of conditions was determined manually in 24-well Linbro plates (Hampton Research) using the hanging drop vapour diffusion method. 1-2 μ l of protein were used in each of the 24-wells, where 500-1000 μ l of crystallization buffer was used in each well. Conditions for optimization were obtained by varying the protein concentration, the precipitant concentration, the pH and the temperature.

2.15.2 Cryoprotection of crystals

Good-looking crystals were picked and soaked in a cryoprotectant buffer, which was used to protect crystals from freezing and from X-ray radiation damage. The cryoprotectant solution was composed of a cryoprotectant agent such as ethylene glycol, glycerol, sugar or low molecular weight PEG, mixed with precipitant. In this thesis cryoprotected crystals were frozen in a nitrogen stream at 100k. Different concentrations of ethylene glycol were used as the cryoprotectant agent. The quality of diffraction from crystals frozen in each cryoprotectant condition was judged on

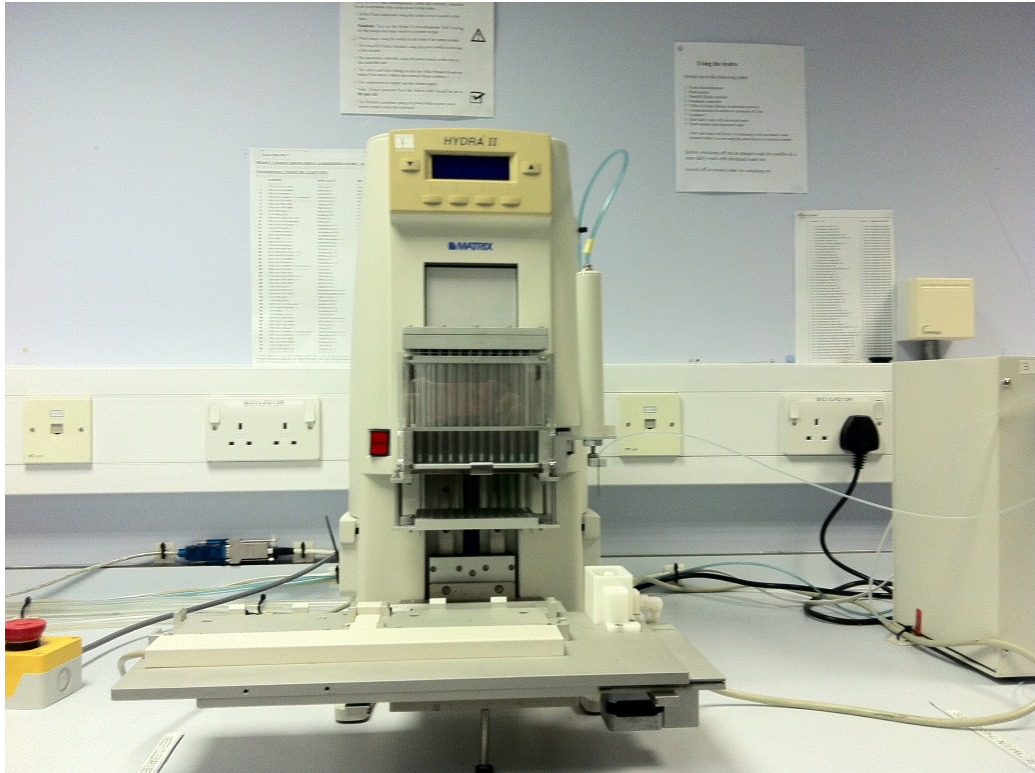


Figure 2.3: The Matrix Hydra II Plus One crystallization robot.

an in-house X-ray source at 100 K using the Rigaku Micromax 007 copper rotating anode generator that produces X-rays at a wavelength of 1.542 Å to choose the optimum cryoprotectant condition. The best cryoprotectant solution is the one that offers a clear appearance and maximizes protein diffraction.

2.15.3 Data collection and processing

The collection of X-ray diffraction data was achieved by mounting crystals between an X-ray source and an X-ray detector. Diffraction images were collected by rotating the crystals in the X-ray path. Initial crystal testing was done using a Rigaku MM007 copper rotating anode generator and a Mar Research image plate detector at the Molecular Biology and Biotechnology department, the University of Sheffield. For generating high quality diffraction data, synchrotron X-ray beams at the Diamond Light Source in Oxford, United Kingdom were used. The collected data were integrated using IMosflm program (Leslie, 1994) at home (Sheffield X-ray generator) or XDS and xia2 (Kabsch, 2010b) at the Diamond Light Source. The XSCALE program (Kabsch, 2010a) was used to scale the integrated data. Data collection for this thesis was carried out at the Diamond light source in Oxford on the I02 and I04 beamlines.

2.15.4 Structure determination

Molecular replacement (with a known structure of high identity) using the program PHASER (McCoy *et al*, 2007) from the CCP4 suite was run to solve the structure. Water and un-modeled molecules of electron density were defined using the program COOT (Emsley & Cowtan, 2004). Refinement of the model against the electron density map was accomplished using the program REFMAC (Murshudov *et al*, 1997). The electron density map was improved and the model was rebuilt between the rounds of refinement using the program COOT. In this work we were able to determine three forms of the BPSL2418 protein structure (BPSL2418-bound MES, BPSL2418-bound Met-SO and reduced BPSL2418), more details on solving these structures are in chapter 5.

2.16 Electron Microscopy

Electron Microscopy was used to investigate the CT2 oligomeric state. Diluted samples of His-CT2 (0.1 mg/ml) were negatively stained using 2% Ammonium Molybdate Stain, pH 7. Negatively stained electron images were recorded on a Philips CM100 transmission electron microscope, by a Gatan MultiScan 794 charge-coupled device camera (Ohi *et al*, 2004).

Chapter 3: Theory of X-Ray Crystallography

Several molecules, including proteins are able to solidify, forming crystals under certain conditions (Rhodes, 2006). X-ray crystallography is an approach that is used to identify the three dimensional structures of biological molecules. In addition to the X-ray technique, the three-dimensional structures can be determined using nuclear magnetic resonance (NMR) and cryo-electron microscopy (Brändén & Tooze, 2009). It is clear that X-ray crystallography is most used procedure as thousands of X-ray structures are saved in the Protein Data Bank (Barman *et al*, 2013).

This chapter gives a brief demonstration of the methods and theory of X-ray crystallography, as it is the main used technique in this thesis. The information in this chapter was mainly provided from these bibliographic sources: 'Crystallography Made Crystal Clear' (Rhodes, 2006), 'Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology' (Rupp, 2010) and 'Protein Crystallography' (Blundell & Johnson, 1976).

3.1 The use of X-rays to produce images of protein molecules

In order for a specimen to diffract light and be observed under magnification, the shortest distance between two separate points must be larger a half of the light wavelength. However, the wavelength of visible light is limited between 400-700 nm, therefore light microscopes cannot obtain an image of molecules smaller a 200 nm. To obtain an image of individual atoms in macromolecules, for example proteins, in which distance between atoms are only about 0.15 nm, electromagnetic waves with shorter wavelength (about 0.1 nm) such as X-rays must be used. X-rays waves are diffracted by individual atoms but cannot be focused by lenses (as visible light waves are in light microscopes) thus an image cannot be viewed directly. Alternatively the X-ray diffraction of the object can be recorded by measuring the directions and strengths (intensities) of the diffracted X-ray beams

using a detector. The measurements are transferred to a computer, takes the place of the action of a lens, to obtain a graphics image of the object. To associate the diffraction pattern to the structure of the object, a mathematical construct called Fourier Transform can be used. However X-ray diffraction from a single molecule is too weak to be detected, but the ability of many molecules including proteins to form crystals can solve this problem, where a crystal is composed of individual molecules arranged in an ordered lattice so that the contribution from individual molecules are added together to give detectable X-ray diffraction.

3.2 The features of crystals

Under certain conditions many macromolecules including proteins are able to solidify forming crystals. In state of protein crystallization, each molecule of the protein appropriate one or a few identical orientations, is produces a crystal in an orderly three-dimensional array of molecules. The molecules in protein crystals are packed together, forming a repeating lattice, by non-covalent interactions and a mixture of protein-protein and water-mediated hydrogen bonds. A crystalline lattice consists of an identical repeating unit, which can include multiple copies of the protein and is called a unit cell; this is the smallest component that is totally representative of the whole crystal (Figure 3.1). The unit cell has three dimensions defined by three lengths a , b and c and three angles α , β and γ . There are seven various crystal systems, which are determined by dimensions and angles of the unit cell: triclinic ($a \neq b \neq c$, $\alpha \neq \beta \neq \gamma \neq 90^\circ$) the simple crystal system, monoclinic ($a \neq b \neq c$, $\alpha = \gamma = 90^\circ$ $\beta \neq 90^\circ$), hexagonal ($a = b \neq c$, $\alpha = \beta = 90^\circ$ $\gamma = 120^\circ$), cubic ($a = b = c$, $\alpha = \beta = \gamma = 90^\circ$), tetragonal ($a = b \neq c$, $\alpha = \beta = \gamma = 90^\circ$), orthorhombic ($a \neq b \neq c$, $\alpha = \beta = \gamma = 90^\circ$) and rhombohedral ($a = b = c$, $\alpha = \beta = \gamma \neq 90^\circ$). The lattice types can be sub-divided into five types depending on the position of the molecules inside the unit cell: primitive, body-centered, face centered and base centered. Within the unit cell the biggest aggregate of molecules, which has no internal crystallography symmetry, but can be superimposed on other exact adjacent aggregates of molecules by applying crystal symmetry operations is defined as the asymmetric unit. The symmetry operations that used to represent unit cell symmetry are rotation, translation, inversion and reflection. According on the crystal system, lattice type and the

asymmetric unit the crystal can be assigned to a space-group. The space-group provides a description of the symmetry of the crystal, where a total of 230 space-groups, can be formed for the non-chiral molecules by applying combinations of symmetry operations, but only 65 space-groups are possible for chiral protein molecules because mirror symmetry cannot be applied. The unit cell dimensions and symmetry of the space-group can be derived from the diffraction image.

3.3 Producing protein crystals

Producing crystals can be a complicated and time consuming step in the X-ray structural analysis of a protein (Drenth, 1999). Crystal formation demands dissolving purified protein in an aqueous buffer including a precipitant. These chemical precipitants are sufficient to control the precipitation of proteins without causing protein denaturation. Several chemical precipitants are able to drive this process, including salts such as ammonium sulfate, high molecular weight straight chain polymers, such as PEGs and organic solvents (Rhodes, 2006). The procedure of protein crystallization driven by a precipitant is a phase development phenomenon (Figure 3.1). Protein crystallization is basically set up with a protein in an aqueous buffer containing precipitant molecules at concentrations that allow protein precipitation. As the procedure progresses, the water in the solution evaporates leading to an increase in the protein and precipitant concentrations. This moves the protein molecules from the bulk of the solution, and the molecules start either to integrate in ordered layers forming crystals, or form amorphous precipitant (Rhodes, 2006). There are three stages of protein crystallization (Figure 3.1), the first stage is the unsaturated phase at which the solution is not saturated with protein. Crystal production appears in the other two stages the nucleation and growth and the growth phases, where the solution is supersaturated with protein. The nucleation is the primary production of molecular clusters from where crystals grow, this phase occurs once the concentration of the protein and/or precipitant is higher than the optimal condition for slow precipitation. When the concentration of protein and precipitant is reduced permitting slow precipitation the crystal growth phase occurs.

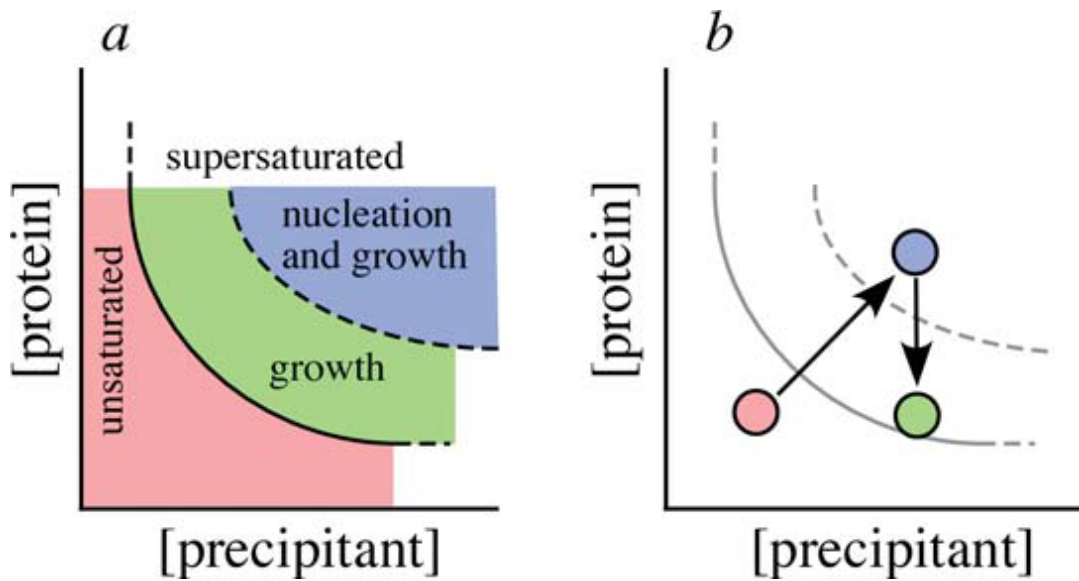


Figure 3.1: Protein crystallization phases. (a) A diagram showing the crystallization phases. The red area represents the unsaturated stage when the solution is not saturated with protein. The blue area represents the conditions that allow both nucleation and growth. The green area represents the conditions that support growth only. (b) An ideal approach for achieving large crystals is the crystallization process moves from the unsaturated phase to the nucleation and growth phase, then move quickly to the growth phase until crystal growth quits. Adapted from (Rhodes, 2006).

The decreased supersaturated level allows individual protein molecules to incorporate in an orderly manner. The rapid precipitation in nucleation stage is more likely to generate a large number of small crystals or amorphous solid but the slow precipitation is more likely to produce a small number of large crystals. Thus in the ideal approach the nucleation phase occur first, then quickly the growth stage start preventing more nucleation and supporting a few larger crystals to be produced.

3.3.1 Crystallization techniques

The initial crystallization hit was obtained using a Matrix Hydra II PlusOne robot as mentioned before in section 2.15.1. In order to produce larger and higher quality crystals the initial conditions can be optimized by changing protein concentration, precipitant concentration, ratio of protein to precipitant in drops, volume of drops, pH and temperature etc. The vapour diffusion method is a manual technique commonly used to optimize the initial screens and to produce desirable crystals. The method requires blending the protein and crystallization solution in a suitable ratio. Then the mixture and a reservoir of a precipitant are placed separately and close to each other in a sealed container and diffusion of volatile solvents occurs between them through the vapour phase. The concentration of the crystallization mixture is less than the concentration of the reservoir that induces the solvent/water to transport from the crystallization mixture to the reservoir. Consequently the concentration of protein and precipitant increase and this pushes the protein out of the solution. There are two types of vapour diffusion method: The hanging drop vapour diffusion and the sitting drop vapour diffusion.

3.3.1.1 Hanging drop technique

In hanging drop technique (Figure 3.2), usually 1 to 2 μl of purified protein at high concentration are combined with crystallization solution (precipitant), often in a 1:1 ratio. The protein/precipitant drop is pipetted onto a clean pre-siliconised glass coverslip, which is inverted and suspended over a well of 500-1000 μl precipitant solution. The coverslip is sealed to the well using immersion oil to ensure an airtight seal.

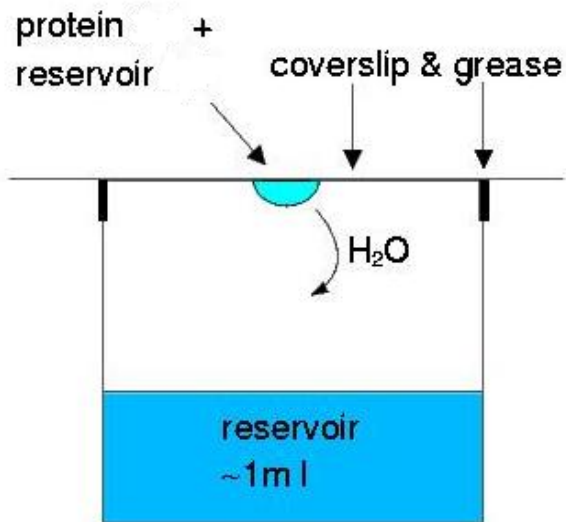


Figure 3.2: The hanging drop vapour diffusion.

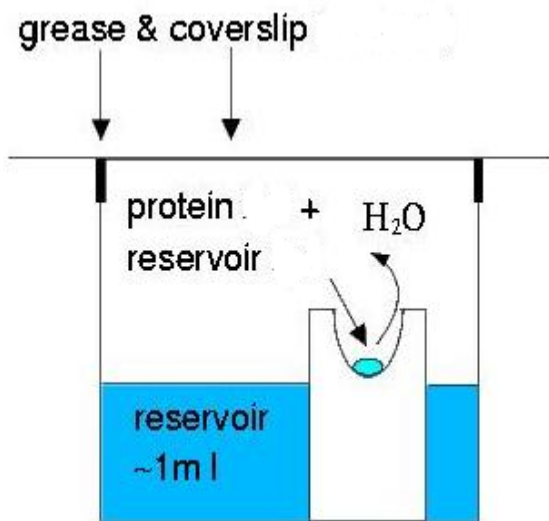


Figure 3.3: The sitting drop vapour diffusion.

3.3.1.2 Sitting drop technique

In a sitting drop method the protein/precipitant drop pipetted in a well of a micro-bridge, which placed above a reservoir of precipitant solution (Figure 3.3). As in the hanging drop the well is sealed using a coverslip and sticky oil. In hanging drop and sitting drop vapour diffusion, usually a 1:1 ratio of protein to precipitant is applied, but changing this ratio can improve crystal size.

3.4 Crystal mounting

In order to collect diffraction data, the protein crystal should be mounted onto a goniometer head, which is a device to control the orientation of the crystal while it is exposed to an X-ray beam. The protein crystal is typically mounted in a loop composed of a twist of nylon or plastic material connected to metal pin ending with magnetic base. The magnetic base facilitates the binding of the loop onto the goniometer head (Figure 3.4). Crystals were mounted with a droplet of the mother liquor plus cryoprotectant agent, to keeps the crystal hydrated and to prevent crystals damage. The crystals are held in the loop by surface tension in a droplet of liquid and permits for crystal transferring without drying them. The good quality crystals produce sharp diffraction patterns at large angles emerging from the X-ray beam.

3.4.1 Cryo-cooling protein crystals

In the early stage of protein crystallography, crystals were mounted at room temperature, but this was associated with crystal radiation damage, which consequently affects the diffracting quality of the crystal. Exposure of protein crystal to X-rays produces free radicals by ionizing X-ray photons, which cause either a localized damage, such as destruction of disulfide bonds or a general damage to the whole crystal. However, the less quality of crystal diffraction means that several crystals will used to collect sufficient data for structure identification. To prevent this trouble, the crystal needs to be soaked in

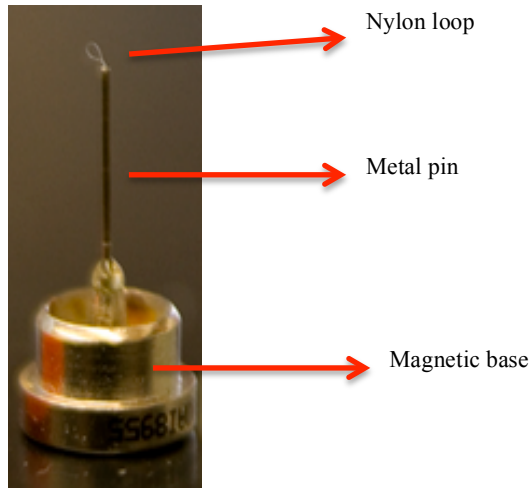


Figure 3.4: This photo shows the loop that holds the crystal. The loop consists of a nylon twist, metal pin and magnetic base. The loop is then mounted on a goniometer head.

cryoprotectants and X-ray data have to be collected at very low temperature. By cooling the crystal to liquid nitrogen temperature, molecular order in the crystal will raise and automatically develop crystal diffraction. Using this advantage enables the collection of complete data sets from single crystals.

3.4.2 The requirement for a cryo-protectant

The technique of mounting crystals in a cryoprotected mother liquor is necessary to maintain the crystal hydrated and to reduce radiation damage to the crystals. This method includes transferring crystal from the drop of mother liquor from where it has grown to a solution that contains mother liquor and convenient cryo-protectant like ethylene glycol, glycerol or low molecular weight polyethylene (PEGs). The presence of cryo-protectant prevents the water inside and surrounding crystal from freezing and forming ice crystals that will also diffract. It is important to optimize the length of soak in cryoprotectant, type of cryoprotectant and cryoprotectant concentration depending on the size of the crystal and a number of experiments that need to be applied.

3.5 Collecting X-ray data

The cryoloop including a crystal is mounted onto a goniometer head between an X-ray source and X-ray detector. The mounted crystal is kept cold by a cold nitrogen gas stream generated from a liquid nitrogen source. The goniometer rotates the crystal in the X-ray beam while a series of diffraction images are collected, from the different degrees of rotation. Usually each image is collected by rotating the crystal through a small angle of about $0.1-1.0^\circ$, and this will produce a dataset which can be used to determine the protein structure. The spacegroup of the crystal and the degree of mosaicity can determine the whole dataset size and the number of degrees of rotation for each diffraction image. For example a smaller dataset is required when the crystal belongs to a higher symmetry space group, as the equivalent reflections will appear more often. Also a small degree of rotation per diffraction image is recommended to use with crystal having a large cell

dimension or a crystal with a high mosaicity to separate the reflections while diffraction images are collected.

3.6 X-ray sources

In X-ray crystallography three sources are used to generate X-rays: Rotating anode tubes (Figure 3.5) and synchrotron radiation source (Figure 3.6), which produce more intense X-rays.

3.6.1 Rotating copper anode tubes

The rotating anode X-ray source is an evacuated tube with a rotating copper anode at one end and a cathode (usually a tungsten filament) on the other end. The X-rays are produced by bombarding a metal target (the rotating anode) with electrons that have been generated from the heated filament. The cathode is heated, which allows the electrons to be released and accelerated towards the rotating copper anode through an electronic field. These high-energy electrons collide with, and displace, low-energy electrons in the target metal. When a high-energy electron hits an electron from a low energy orbital, the electron from higher energy orbital will drop down, emitting an X-ray photon. The wavelength of the produced X-ray radiation is determined by the metal that X-rays emits from. For protein crystallography, the anode is usually made of copper due to its high heat conductivity. The copper anode is able to emit two high-energy X-rays: $K\alpha$ has a wavelength of 1.54 Å while $K\beta$ has a wavelength of 1.39 Å. The produced X-ray radiation emerges from the rotating copper anode tube through windows of beryllium. Initial crystal data collection for this thesis, was performed at the crystallography lab, in the University Of Sheffield, using a Rigaku Micromax 007 copper rotating anode generator (Figure 3.5).



Figure 3.5: A Rigaku Micromax 007 copper rotating anode generator.

used in X-ray crystallography lab in MBB/ The University of Sheffield. X-rays are generated from a rotating copper anode source (not shown), and come through a metal collimator (1). A cryo-loop containing the crystal is mounted on the goniometer to be struck by the X-ray beam (2). An image plate detector captures the X-ray reflections (3). A nitrogen gas source to keep the crystal at cryogenic temperature.

3.6.2 Synchrotron

A synchrotron X-ray source (Figure 3.6) is a big particle accelerator that is able to generate electrons and accelerate them nearly to the speed of light to produce higher intensity X-rays with wavelengths in the range from 0.5 Å to 1.8 Å. A synchrotron contains a particle accelerator and a circular path around a central storage ring and a series of bending magnets. Electrons are injected by a particle accelerator inside the ring and accelerated close to the speed of light controlled by radio transmitters and kept in a circular motion due to the powerful magnets (Figure 3.7). The obligatory curved path leads the electrons to lose energy and to change velocity, which cause the emission of X-ray radiation at tangents to the circle (beamlines) (Figure 3.7). The synchrotron has also two accessory devices; wigglers and undulators, which can increase the intensity of the radiation by changing the direction of the electrons several times during a short distance. A series of beamlines are located tangentially around the storage ring supplying X-ray beams to where crystals are mounted for data collection. The synchrotron has many distinct features, such as the high intensity of radiation that permits collecting X-ray data in shorter exposure times before the protein crystals are damaged, and also allows data collection from very small crystals. Further, the capability of selecting the wavelength of the X-ray beam is a very useful feature of the synchrotron for phase determination using Single Wavelength Anomalous Dispersion (SAD) or Multi Wavelength Anomalous Dispersion (MAD).

3.7 X-ray Detectors

X-ray detectors are able to record the intensity and direction of the crystal diffracted X-rays. A pattern of crystal X-ray diffraction is obtained through the use of detectors.

3.7.1 Image plate detectors

The image plate detector is made up of a plate of plastic coated in thin sheet of



Figure 3.6: An aerial view of Diamond Light Source, the UK's national synchrotron science facility, at the Harwell Science and Innovation Campus near Didcot, Oxfordshire. Adapted from the DLS website.

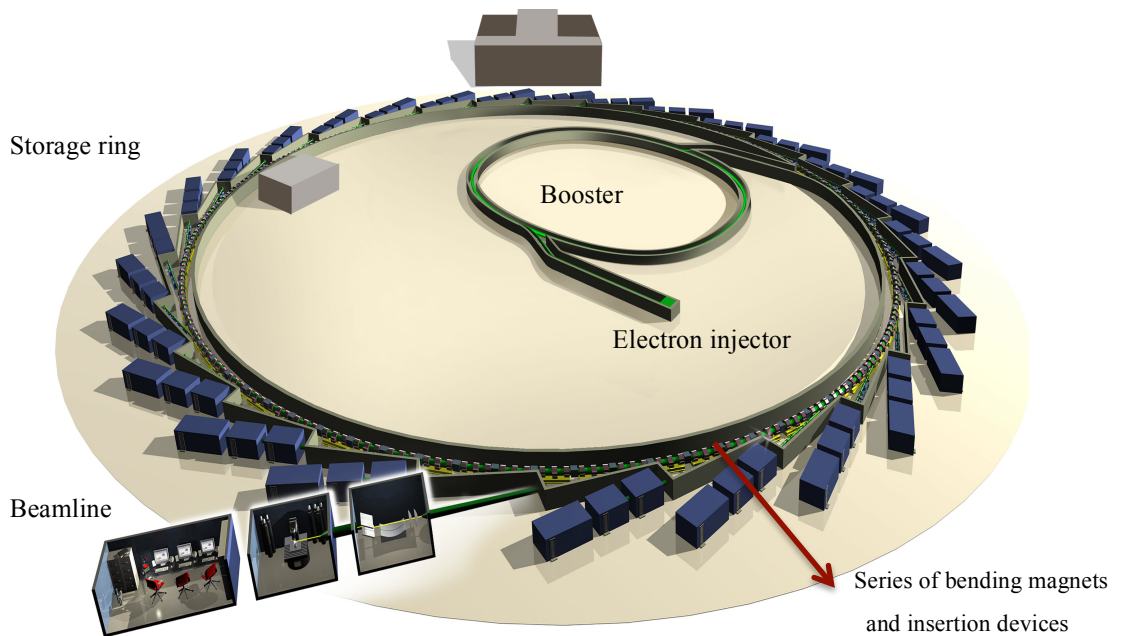


Figure 3.7: Schematic drawing of Diamond Light Source. Adapted from the DLS website.

phosphorescent material, usually crystalline BaFEu. Once X-ray photons hit the plate they interact with the Eu^{2+} leading to excitation of the phosphorescent material. In order to record the X-ray, a He-Ne laser scans the plate leading to emission of a blue light at a wavelength of 390 nm that is captured by a photomultiplier and transformed to a digital signal. A high intensity tungsten light is used to erase the detector and return the phosphorescent material into a ground state. The image plate detector takes a long time to collect a data set, which can be a disadvantage of using this type of detector. Initial crystal testing for this thesis was run at Sheffield in the crystallography lab using MarResearch MAR345 image plate detector.

3.7.2 Charged-coupled device (CCD) detectors

In a charged-coupled device detector the image is immediately appeared to the user in a fluorescent screen linked to a charged-coupled device by fiber optical strands. Crystal diffracted X-rays hit the phosphor screen and are transformed into photons of visible light, which are conducted towards the CCD chips over the fiber optical strands. The total of signal that knock the CCD chips corresponds to the total X-ray photons hitting the detector. The detector systems on the diamond synchrotron lightsource are the charged-coupled device detector. The advantage of the CCD detectors is the reduction of data collection time, which consequently reduces the amount of radiation damage, can affect the crystal.

3.7.3 Pilatus Detectors

The Pilatus system is a single photon counting pixel array where the incoming photons are directly measured allowing data to be collected continuously. The principle advantage of the Pilatus system over CCD detectors, and CCD detectors over image plate detectors, is the reduced read-out time, reducing the data collection experiment time and therefore the amount of radiation damage sustained by the sample. The Pilatus detector can also produce a greater signal to noise ratio by applying an energy threshold to the photons that are counted.

3.8 The principles of X-ray diffraction

3.8.1 Diffraction from crystals

During the procedure of shooting X-rays through crystals to generate diffraction, most of the X-rays do not clash with the crystal and directly pass through, colliding with the backstop. Occasionally an X-ray hits an electron surrounding an atom inside the crystal, leading to electron oscillation. The electron oscillates at the same frequency as the incident X-ray generating a secondary wave, which is recognized by the detector as a reflection. These reflections are representative of the crystal on a particular set of lattice planes, which slice into the crystal, determined by Miller Indices, but only if Bragg's law is satisfied.

3.8.2 Bragg's Law

A crystalline state is composed of atoms arranged in repeated ordered arrays forming a lattice. The crystal unit cell contains planes that divide up and cut across the axes of the unit cell, which determined by Miller index $h k l$. Bragg's law revealed that a set of parallel planes with Miller index $h k l$ and different interplanar spacing $d_{h k l}$ generates a reflected beam when X-rays of wavelength λ strike upon the planes at an angle θ and reflected at the same angle, when θ reaches the condition. Bragg's law describes the conditions necessary for a set of lattice planes to generate a reflection.

If n is an integer and λ is the wavelength of the X-ray that hit the planes (incoming X-rays); d is the distance between planes with indices $h k l$; θ is the angle of the incidence of the incoming X-rays, which strike the planes. The diffracted waves are in phase and will contribute to a reflection spot if the difference in X-rays path length (the distance $2d\sin \theta$) is equal to an integral number of the wavelength of the incoming X-rays (Figure 3.8) and so:

$$n \lambda = 2d \sin \theta$$

If Bragg's law is not satisfied, the X-ray diffraction will not be observed and not produce a reflection and this will give a diffraction image of mainly blank space.

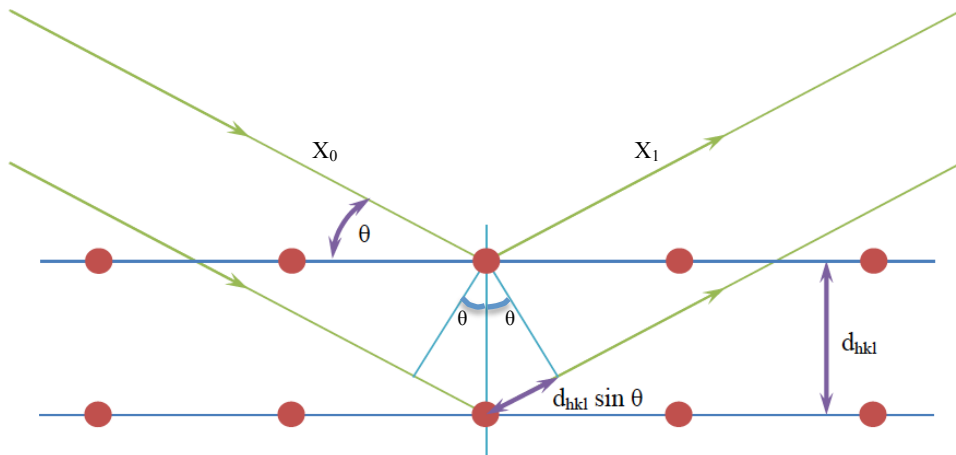


Figure 3.8: Bragg's law describes the needed condition for a set of lattice planes to generate a reflection. X_0 is the incoming X-ray beam and X_1 is the diffracted beam (outgoing beam) which hits the detector. d is the distance in angstroms (\AA) between the lattice planes. The diffracted waves X_1 will be observed only if the distance $2d\sin \theta$ is equal to an integral number of the wavelength of the incoming X-rays.

3.9 From diffraction to electron density

The collected data is made up of a list of intensities of reflections each with an index $h k l$ representing a position in reciprocal space and that refer to atoms on a set of lattice planes in real space. Each reflection in a diffraction image includes contribution from all atoms in the unit cell and it can be described by a structure factor (F_{hkl}).

$$F_{hkl} = \sum_{j=1}^n f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

The structure factor (F_{hkl}) is a Fourier sum of the contributions of each atom (j) within the unit cell containing (n) atoms, to the reflection with indices h, k, l in the reciprocal lattice. (f_j) is the scattering factor of each atom (j) and it provides the amplitude of the wave contribution of atom (j) to the Fourier sum. The contribution of each atom to the Fourier sum is treated as a sphere of electron density dependent on the number of electrons present in the atom. The exponential term represents a simple three-dimensional periodic function with both sine and cosine components. It represents the position of the atom (j) within the unit cell as (x_j, y_j, z_j) and h, k, l are the indices of the specific reflection in reciprocal space. Alternatively (F_{hkl}) can be defined by the electron density volume element (ρ) of each atom:

$$F_{hkl} = \int_V \rho(x, y, z) e^{2\pi i(hx + ky + lz)} dv$$

$\rho(x, y, z)$ is the electron density at position (x, y, z) and (V) is the volume of the unit cell and the integral over V reveals that the integration includes all values of x, y, z .

3.10 Processing data from diffraction images

3.10.1 Data indexing and integration

When a dataset is collected, the measurement of the intensity of each reflection will be needed. A dataset was automatically indexed and integrated using programs such as MOSFLM and XDS. For data indexing these programs are able to measure the distance between reflections to determine the dimension of the unit cell and the space group that the crystal belongs to. They can measure the intensity of each reflection (spot) in pixels in contrast to the background level of the detector and combine this value with the information of the spot position (hkl). The data are then integrated where all collected images containing all reflections are converted into a single file (the mtz file).

3.10.2 Data merging and scaling

The program SCALA was used to merge and scale symmetry related reflections across all images so that they have a consistent intensity scale. Scaling images must be from the start to the end of the dataset and from low to high resolution on each image. The quality of the dataset can be gauged by the R-factor and R_{merge} , which estimate the agreement of related reflections and the correlation within the dataset. The repetition of the same reflection or the degree of multiplicity can be measured through the series of dataset by the R_{pim} and $CC1/2$, which also can be used with R_{merge} to evaluate the quality of dataset. R_{pim} is an alternative measure that also takes account of multiplicity. $CC1/2$ has superior properties compared with R_{merge} . Its properties in the presence of systematic error are documented.

3.11 structural determination

3.11.1 initial phasing

During the diffraction experiment the intensity of each reflection is only measured while the phase cannot be directly recorded, this known as the phase problem. Therefore the initial phase needs to be obtained by other methods such as the

multiple-wavelength anomalous diffraction (MAD), single-wavelength anomalous diffraction (SAD) and molecular replacement (MR) experiments. In the BPSL2418 project the method used to estimate the phases is the molecular replacement (MR).

3.11.1.1 Molecular replacement (MR)

In this method protein homolog with high sequence identity to the target protein can be used as a search model to determine the orientation and position of the molecules inside the unit cell. The obtained phases can be used to produce electron density maps. The existence of high sequence identity homologs to BPSL2418 allows the use of molecular replacement method in this project. MR can be run using a program called PHASER (McCoy et al, 2007). The presence of a homologue protein permits for applying a calculated set of phases to the experimental intensities to find a solution for the phase problem. In order to do this we must know the rotational and translational position of the protein molecules within the unit cell, which can be obtained by comparing the Patterson Function of the search model and the trial model.

A Patterson Function is a vector map that shows the atomic distances, which are correlated with the crystallized molecule. Molecular replacement demands first that the search model to be rotated in all possible orientation as the unknown protein in the unit cell. This can be done by comparing Patterson map of the search model and the unknown protein. The next step is the translation function, which is required to determine the correct position within the unit cell.

The rotation function includes comparisons between the search model (Known protein) Patterson map and the unknown protein Patterson map in different orientations. The Patterson maps are aligned and oriented in steps through three dimensions. In each orientation step the structure factor amplitudes of the trial model were calculated and compared with the known model structure factor amplitudes. The rotation function solution is obtained when the maximum overlap or similar orientation is found. Then this solution is applied to the search model's coordinates and new rotated model is put out and used in the translation function step. The translation function is a positioning process of the protein to be placed in the correct orientation and position in the asymmetric unit. The rotated model

demands be translated into the correct coordinates, x,y,z axis in order to be placed correctly within the unit cell. A similar comparison is made between the rotated model Patterson map and the unknown model Patterson map to find the best solution. The electron density map can be produced by calculating the phases from the search model by Fourier Transform and are combined with the experimentally derived intensities.

3.12 Structure re-building and refinement

Once the molecular replacement obtains the suitable model, this model needs to be refined and improved. The program COOT was used to visualize the electron density map to re-build the model. The REFMAC5 program in the CCP4 suite was used to refine the model by running several cycles of refinement and re-building to boost the agreement between the model and the electron density. The refinement includes repositioning side-chains, mutating residues, defines density aspect and modeling ligand, adding water molecules. The refinement progressing can be evaluated by comparing the value of the R-factor and the free R-factor. The R-factor (reliability factor) is a measure of the agreement between the obtained model and the collected data.

3.13 Structure validation

During the refinement cycles it is important to monitor the difference between the structure factors amplitudes of the observed data (the experimental data) F_{obs} and the calculated search model F_{calc} . The quality of the refined model is associated with the lower value of R-factor. The R-factor measures the different between F_{obs} and F_{calc} and it is described as:

$$R = \frac{\sum_{hkl} \left| |F_{obs}| - |F_{calc}| \right|}{\sum_{hkl} |F_{calc}|}$$

During scaling a test set of the observed data usually 5% is excluded from the refinement process (R-free), which is described as:

$$R_{free} = \frac{\sum_{hklCT} \left| |F_{obs}| - |F_{calc}| \right|}{\sum_{hklCT} |F_{obs}|}$$

If R-factor and R-free reduce during the refinement and rebuilding of the model this indicates correct changes have been made to the model.

After refinement the final atomic coordinates should be validated by checking that the model represents the targeted protein very well. This includes the analysis of a Ramachandran plot, the geometry of main and side chains, clashes between atoms. This can be made by using the PROCHECK (Laskowski, 1993) and the Molprobit (Chen *et al*, 2010) server.

Chapter 4: BPSL2418 Cloning, Expression, Purification and Crystallization

This chapter represents the cloning, overexpression, purification and crystallization for *Burkholderia pseudomallei* BPSL2418 the MES-bound, substrate-bound, unbound and reduced forms.

4.1 Cloning of BPSL2418

4.1.1 Amplification of the *bpsl2418* gene

The *bpsl2418* gene was amplified from *Burkholderia pseudomallei* strain D286 genomic DNA by PCR as described in section 2.4., using the BPSL2418 forward and reverse primers and BioMix Red reaction mixture. To analyze the PCR products, they were run on a 1% agarose gel (see section 2.5). The agarose gel shows a band of the expected size of approximately 513 bp (Figure 4.1), and the Gel Extraction Kit was used to purify the PCR products (see section 2.6).

4.1.2 Ligation of *bpsl2418* gene into pETBlue-1 vector and transformation into Novablue competent *E. coli* cells

The insert was ligated into the *E. coli* linearized expression vector pETBlue-1 and then the recombinant plasmid was transformed into Novablue cells (see section 2.8.2), which have blue/white screening capability. The ligation with the insert should produce white colonies while the ligation without the insert should produce blue colonies. 4:1 was the ratio of the white colonies to the blue colonies (Figure 4.2). A number of the desired white colonies were picked to inoculate 5 ml LB media each, and grown overnight. Plasmids were extracted from each cell pellet as described in the protocol (see section 2.7).

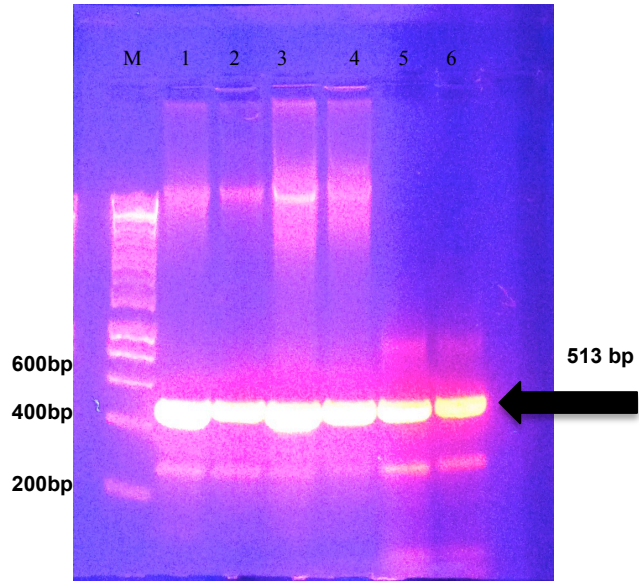


Figure 4.1: 1% DNA agarose gel of *bps/2418* gene amplification: Lane M is DNA hyper ladder 1 molecular weight marker; Lanes 1,2,3,4,5 and 6 are BPSL2418 PCR products.

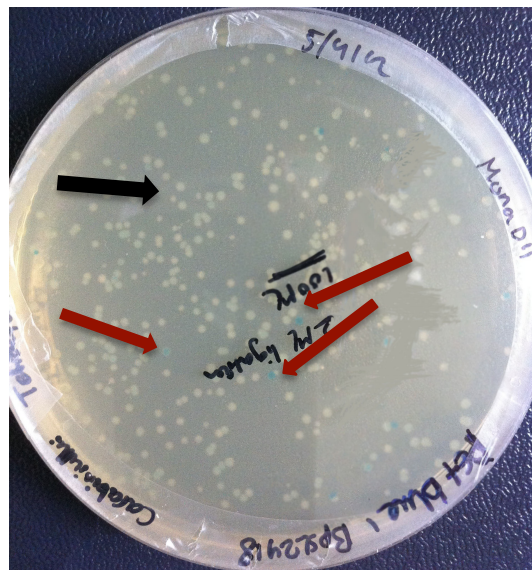


Figure 4.2: Blue/white screening of Transformation of ligation products into competent *E.coli* cells: Black arrow indicates white colonies (ligation with insert). Red arrows indicate blue colonies (ligation without insert).

4.1.3 Confirmation of the presence and orientation of the insert gene in the pETBlue-1 plasmid

The presence and orientation of the *bpsl2418* insert gene in the pETBlue-1 plasmid was confirmed by DNA sequence analysis (Figure 4.3) and by using PCR to amplify *bpsl2418* DNA sequence (513 bp) from the recombinant pETBlue-1 (see section 2.8.3). The PCR reaction produced a band of approximately 523 bp, where the additional 20 bp is the size of pETBlue up primer. The PCR product was visualised on a DNA agarose gel (Figure 4.4).

4.2 Transformation of Tuner (DE3)pLacI competent cells with pETBlue-1 *bpsl2418* recombinants

Cloning in pETBlue vectors regularly requires the use of NovaBlue and Tuner (DE3)pLacI competent cells. NovaBlue allows the blue/white-screening technique that detects the recombinant colonies but contains no source of T7 RNA polymerase. The Tuner (DE3)pLacI is an expression host that provides inducible expression of T7 RNA polymerase. Eppendorf tubes containing 20 µl aliquots of Tuner (DE3)pLacI cells were removed from the -80 °C freezer and incubated on ice for 5 minutes. Once defrosted 1 µl of plasmid DNA (approximately 10 ng/µl) was added and the reaction was left to incubate on ice for 5 minutes. The cells were subjected to a heat shock at 42 °C for 30 seconds before being returned to ice for 2 minutes, and then an 80 µl of SOC media was added to the mixture. The cells were incubated at 37 °C at 250 rpm for 60 minutes before plating 50 µl on LB agar containing 1% glucose, 50 µg/ml carbenicillin and 34 µg/ml chloramphenicol.

4.3 BPSL2418 overexpression

Different small-scale overexpression trials were accomplished for BPSL2418 to obtain the best overexpression condition, and finally the following protocol was used: One colony from transformation plate of Tuner (DE3)pLacI with pETBlue-1 *bpsl2418* recombinant inoculates 3 ml of LB media containing 1% glucose,

```

Query 1   ATGTTTCGCGCTTTCCGAAGCTCTCCCGTCATCGAAGCCCGCGCTCTACGAAACGCTCGCC 60
Sbjct 1   ATGTTTCGCGCTTTCCGAAGCTCTCCCGTCATCGAAGCCCGCGCTCTACGAAACGCTCGCC 60

Query 61  GCTCAGGCACGCGCGCTCGTCGAAACGGAGACCACATCGTCGCCAATGCGGGCAATTTTC 120
Sbjct 61  GCTCAGGCACGCGCGCTCGTCGAAACGGAGACCACATCGTCGCCAATGCGGGCAATTTTC 120

Query 121 GCATCGCTCGTCTATCACTCGCTCGATGGCCTCAACTGGGCCGGGTTCTATTTCTTCGAC 180
Sbjct 121 GCATCGCTCGTCTATCACTCGCTCGATGGCCTCAACTGGGCCGGGTTCTATTTCTTCGAC 180

Query 181 GGTTCGCGAGCTCGTCGTTCGGGCCGTTCCAGGGCAAGCCCGCGTGCCTGCGCATCCCGCTC 240
Sbjct 181 GGTTCGCGAGCTCGTCGTTCGGGCCGTTCCAGGGCAAGCCCGCGTGCCTGCGCATCCCGCTC 240

Query 241 GGCAAGGGCGTGTGCGGCACCGCCGCGCAGACGCGCGCACGAGGTCGTGCACGACGTG 300
Sbjct 241 GGCAAGGGCGTGTGCGGCACCGCCGCGCAGACGCGCGCACGAGGTCGTGCACGACGTG 300

Query 301 CACGCGTTTGCCGGCCACATTGCCTGCGATTTCGGCGTCGCAATCGGAAATCGTCGTGCCG 360
Sbjct 301 CACGCGTTTGCCGGCCACATTGCCTGCGATTTCGGCGTCGCAATCGGAAATCGTCGTGCCG 360

Query 361 CTCGTCGCGCGCGACGGCGCGCTGATCGGCGTATGGGACGTCGATAGCCCGCTCGTCGGC 420
Sbjct 361 CTCGTCGCGCGCGACGGCGCGCTGATCGGCGTATGGGACGTCGATAGCCCGCTCGTCGGC 420

Query 421 CGCTTCGACGCAGNGGATGCGATAGGGATGGAGNNGCTGTGCCGGGTGTTTCGTTCGAGGTT 480
Sbjct 421 CGCTTCGACGCAGAGGATGCGATAGGGATGGAGGCGCTGTGCCGGGTGTTTCGTTCGAGGTT 480

Query 481 GCGTGGGAAAACGCAACGCGGGGNNCCGAGTAA 513
Sbjct 481 GCGTGGGAAAACGCAACGCGGGGCGGCGGAGTAA 513

```

Figure 4.3: The *bpsI2418* insert gene was sequenced and then analyzed by BLAST, <http://blast.ncbi.nlm.nih.gov/Blast.cgi>. BLAST shows that the *bpsI2418* insert gene sequence is 99% identical to hypothetical protein BPSL2418 [*Burkholderia pseudomallei* K96243], with accession number YP_109010. N: nil

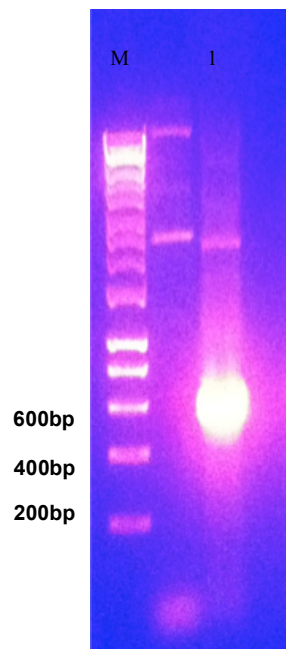


Figure 4.4: Confirmation of the presence and orientation of the insert gene in the pETBlue-1 plasmid: Lane M: hyper ladder 1 molecular weight marker; Lane 1: the PCR product band at the expected size of 523 bp.

50 µg/ml carbenicillin and 34 µg/ml chloramphenicol. The 3 ml primary culture was incubated at 37 °C with shaking at 250 rpm until the OD₆₀₀ reached approximately 0.6-1, then it was stored overnight at 4 °C. The cells were collected the following morning by centrifugation and resuspended in 3 ml fresh media containing the desired antibiotics and 1% glucose. A 100 ml (x10) culture of LB containing all required antibiotics and 1% glucose was inoculated with the 3 ml (x10) starter culture and incubated at 37 °C with shaking at 250 rpm until the OD₆₀₀ reached 0.5-1. Expression was induced by addition of 300 µM IPTG for 4 hours at 37 °C, 250 rpm. BPSL2418 was overexpressed and an SDS-PAGE gel (Figure 4.5) reveals a strong expression band in the soluble fraction. The molecular weight of expressed protein is the same as expected, about 18 kDa.

4.4 Purification of BPSL2418

About 4 g cells were defrosted and suspended in about 35 ml of (50 mM tris-HCl pH 8.0) buffer. Cells were disrupted by sonication and cell debris was removed by centrifugation at 70000 g for 10 min. The protein concentration was estimated in the supernatant fraction (Cell Free Extract) CFE as 7.5 mg/ml. Total protein in the cell free extract was 260 mg. This sample was applied on a 10 ml column with DEAE fast flow Sepharose (2 x 5 ml Hi Trap DEAE Fast Flow cartridges, GE Healthcare) which is a weak anion exchanger. Proteins were eluted with 100 ml gradient of NaCl concentration from 0 to 0.5 M in 50 mM tris-HCl pH 8.0 buffer at flow rate 5 ml/min (Figure 4.6). Four ml fractions were collected and analysed on 12% SDS-PAGE (Figure 4.7). The most pure fractions, 17 and 18, were combined, V=8 ml, C=6 mg/ml, total 48 mg (Peak). Fractions 16, 19 and 20 were also combined, V=12 ml, C=4 mg/ml, total 48 mg (sides).

The peak fractions were diluted 3 fold with ultra pure water and applied to a 6 ml Resource Q column (GE Healthcare), a strong anion exchanger. Proteins were eluted from the column by 90 ml gradient of NaCl from 50 mM to 150 mM in 50 mM MES-NaOH buffer pH 6.3 at flow rate of 4 ml/min (Figure 4.8). 2.5 ml fractions were collected and fractions across the peak were analyzed on a 12% SDS-PAGE (Figure 4.9).

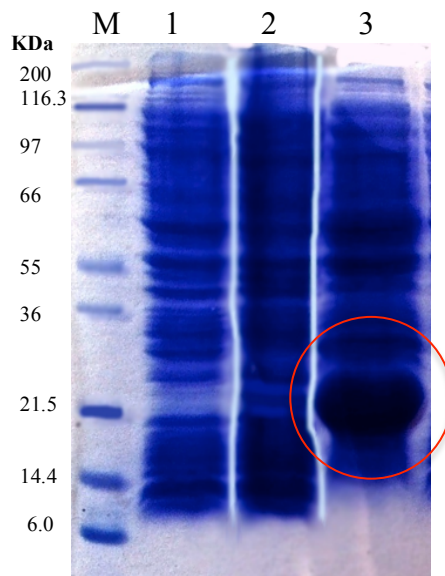


Figure 4.5: 12% SDS-PAGE gel showing the overexpression of BPSL2418. Lane M: molecular weight marker; Lane 1: pre-induction fraction; Lane 2: insoluble fraction; Lane 3: soluble fraction shows a strong overexpression band around 18 kDa.

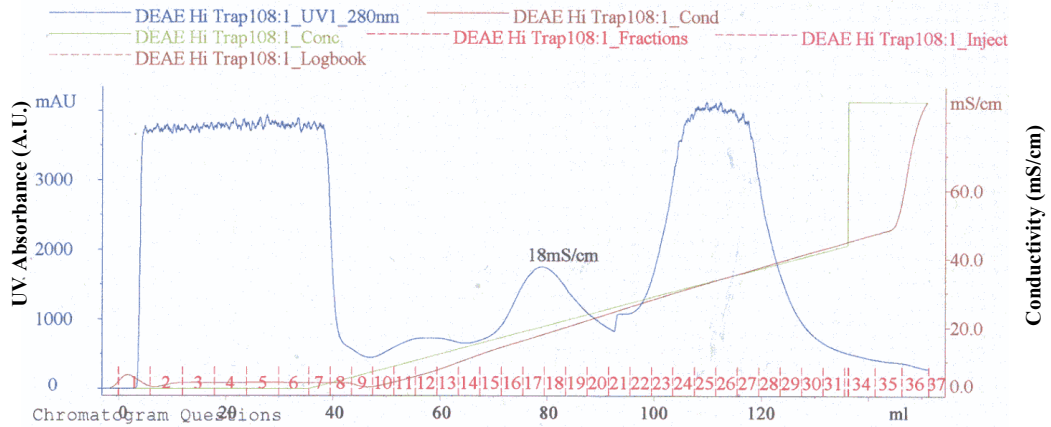


Figure 4.6: A DEAE purification step showing column loading and elution of BPSL2418. This Figure represents the elution of protein with 100ml gradient of NaCl concentration from 0 to 0.5 M in 50 mM tris-HCl pH 8.0 buffer at flow rate 5 ml/min.

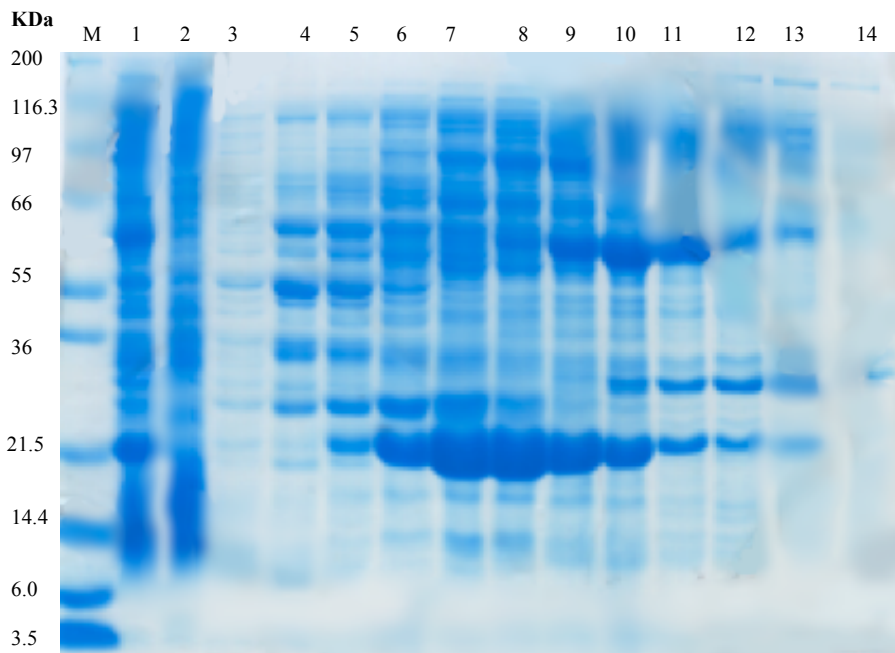


Figure 4.7: 12% SDS-PAGE analysis of DEAE fast flow Sepharose of BPSL2418. Visualised by staining with coomassie blue stain. Lane M: molecular weight marker; Lane 1: CFE; Lane 2: unbound proteins; Lanes 3,4,5,6,7,8,9,10,11,12,13 and 14 are fractions 12,14,15,16,17,18,19,20,21,22,23 and 25 respectively. Fractions 16,17,18,19 and 20 were collected.

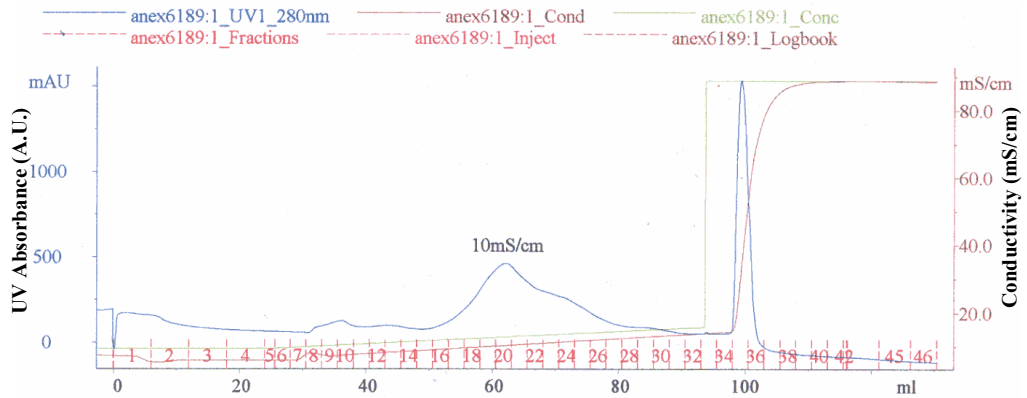


Figure 4.8: A Resource Q column purification step showing column loading and elution of BPSL2418. Proteins were eluted from the column by 90 ml gradient of NaCl from 50 mM to 150 mM in 50 mM MES-NaOH buffer pH 6.3 at flow rate 4 ml/min.

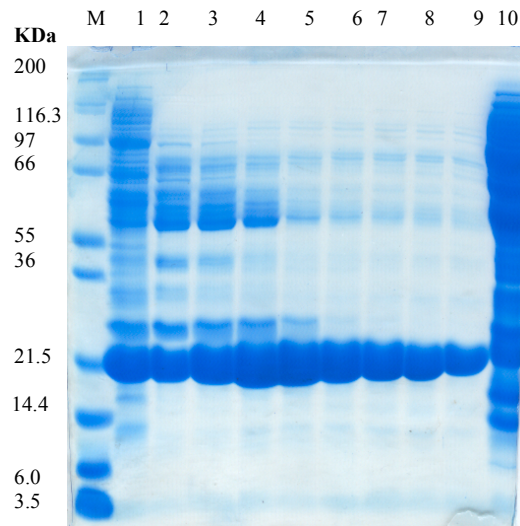
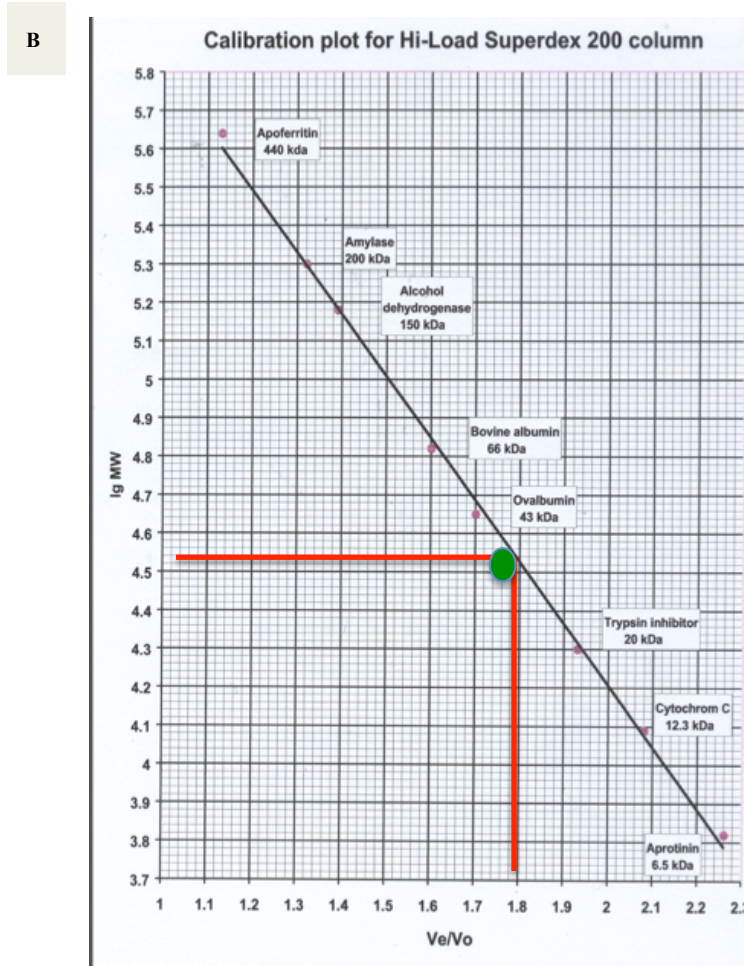
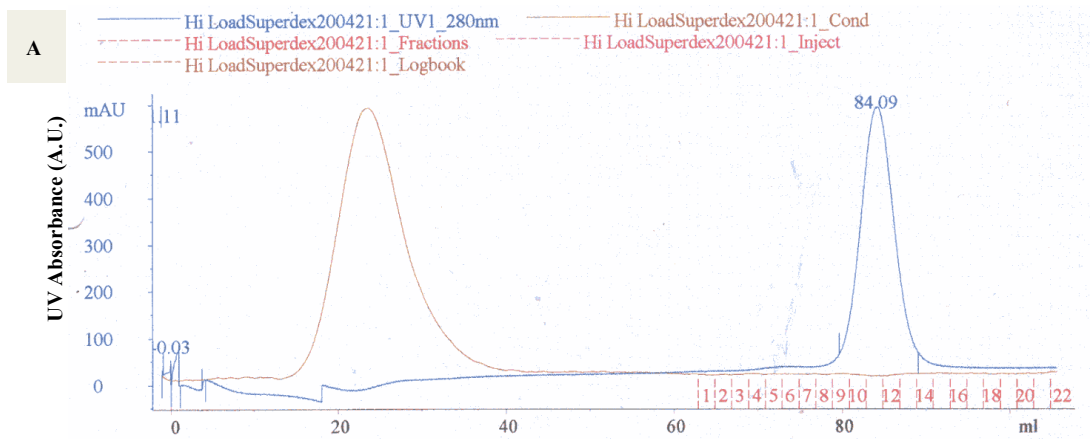


Figure 4.9: 12% SDS-PAGE analysis of Resource Q column purification of BPSL2418. Visualised by staining with coomassie blue stain. Lane M: molecular weight marker; Lane 1: peak fraction DEAE FF; lanes 2,3,4,5,6,7,8,9 and 10 are fractions 18,19,20,21,22,23,24,25 and 35 respectively. Fractions 21,22,23,24 and 25 were collected.

The most pure fractions, 21-25, were combined, $V=12.5$ ml, $C=0.95$ mg/ml, total 12 mg. To prepare a sample for gel filtration, the volume was reduced to 1 ml using a VivaSpin 20 MWCO 10000 concentrator (Sartorius). Fractions 18-20 with less pure protein were also combined, $V=7.5$ ml, $C=1.4$ mg/ml, total 8.2 mg. 7.5 ml of 4 M ammonium sulphate was added to this sample to precipitate it for storage.

The concentrated sample of the most pure BPSL2418 was applied to a 16x60 HiLoad Superdex200 gel filtration column (GE Healthcare). Gel filtration was performed in a buffer containing 0.5 M NaCl 50 mM tris-HCl pH 8.0 at flow rate of 1.5 ml/min. 2 ml fractions were collected starting from 0.5 column volume (Figure 4.10). BPSL2418 was eluted from the column at 84 ml, corresponding to an apparent MW of 38 kDa, indicating its dimeric state in solution. Peak fractions 11 and 12 were combined, $V=4$ ml, $C=1.1$ mg/ml, total 4.4 mg. This sample was concentrated as above and the buffer was changed to 10 mM tris-HCl pH 8.0 using a diafiltration cup. $V=0.2$ ml, $C=20$ mg/ml, total 4 mg. The purification progress of BPSL2418 preparation 1 was analyzed by using SDS-PAGE (Figure 4.11). This preparation 1 was used to screen for crystallisation conditions.

In preparation 2 of BPSL2418 the suspension of precipitated protein from fractions 16,19 and 20 from the DEAE column and fractions 18-20 from the Resource Q column were combined and the pellet collected by centrifugation at 70000 g for 5 min. These stored fractions from preparation 1 are less pure, which suggests the need for an additional purification step to be applied. A hydrophobic chromatography column was also used in preparation 2 of BPSL2418. The pellet was dissolved in 10 ml of 50 mM tris-HCl buffer pH 8.0. $V=10$ ml, $C=2$ mg/ml, 20 mg. The ammonium sulphate concentration in the solution was checked by refraction and was about 0.8 M. The sample was applied to a 5 ml Hi Trap Phenyl-HP cartridge and eluted with an optimised gradient of ammonium sulphate concentration from 0.7 M to 0 M in 50 ml (Figure 4.12). Peak fractions 15-19 were combined, $V=12.5$ ml, $C=0.55$ mg/ml, 6.8 mg. Proteins were precipitated with 2 M ammonium sulphate. The pellet was spun down as above and dissolved in 1.5 ml of 50 mM tris-HCl buffer pH 8.0. The resulting sample, $V=1.9$ ml, $C=5$ mg/ml, 9.5 mg was applied on a gel filtration column. Gel filtration was performed as described above. 12% SDS-PAGE was used to analyse the progress of BPSL2418 preparation 2 (Figure 4.13). Fractions 11-13 from the gel filtration column were



Use the calibration plot for Hi-Load superdex 200 column chart to calculate the protein MW:

$$Ve/Vo = 84/45 = 1.8$$

$$\log MW = 4.51$$

$$MW = 38 \text{ kDa}$$

Note:
 Ve: Elution volume
 Vo: Void volume

Figure 4.10: Gel filtration purification of BPSL2418. (A) a 16x60 HiLoad Superdex200 gel filtration column loading and elution of BPSL2418. (B) the calibration plot for Hi-Load superdex 200 column chart shows that the BPSL2418 corresponding to apparent MW 38 kDa, indicating its dimeric state in solution.

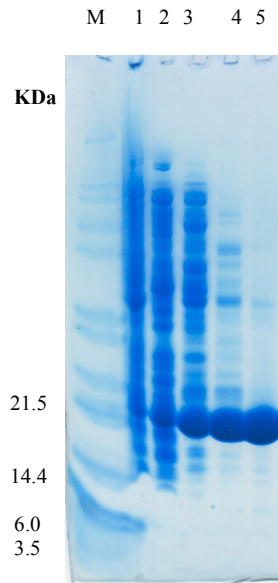


Figure 4.11: 12% SDS-PAGE analysis of purification progress of BPSL2418 preparation 1. Lane M: molecular weight marker; Lane 1: cell debris; Lane 3: CFE; Lane 4: after Resource Q column; Lane 5: after gel filtration.

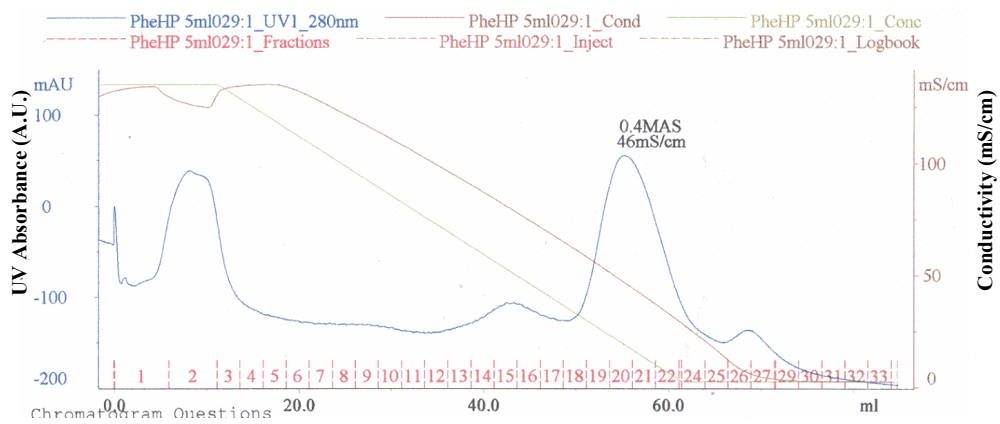


Figure 4.12: A 5ml Hi Trap Phenyl-HP cartridge loading and elution of BPSL2418. The protein was eluted with optimised gradient of ammonium sulphate concentration from 0.7 M to 0 M in 50 ml.

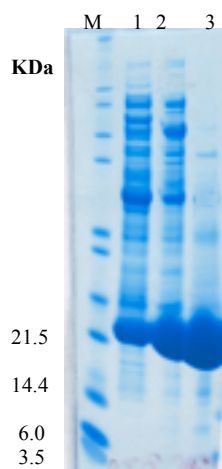


Figure 4.13: 12% SDS-PAGE analysis of purification progress of BPSL2418 preparation 2. Lane M: molecular weight marker; Lane 1: sample applied on Phenyl-HP; Lane 2: after Phenyl-HP; Lane 3: after gel filtration.

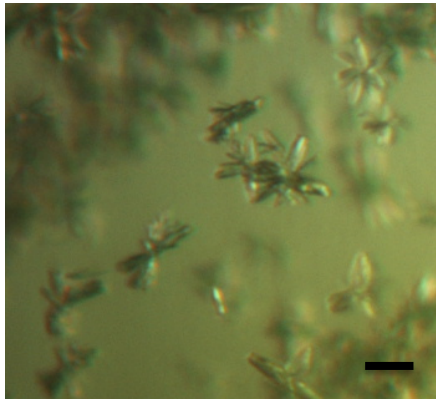
combined, $V=6$ ml, $C=0.9$ mg/ml, 5.4 mg/ml. The sample was concentrated to 30 mg/ml and buffer exchanged as described above. This sample, preparation 2, was used for optimisation of the crystallisation conditions.

4.5 Crystallization trials of BPSL2418

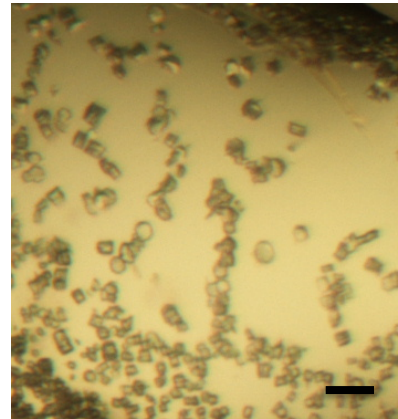
In the first initial trials of BPSL2418 crystallization I found that most of the successful trials contained 2-(N-morpholine) ethanesulfonic acid (MES) buffer. However, as BPSL2418 is predicted to function as a methionine sulfoxide reductase and a MES molecule contains a sulfonic acid group. The formation of crystals of BPSL2418 in complex with MES buffer was encouraging.

4.5.1 Crystallization of the BPSL2418 MES-bound form (BPSL2418_{MES})

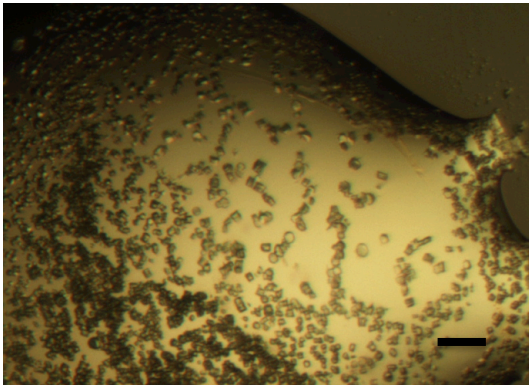
Six initial 96 condition robot screens, the PACT, JCSG, Classics, PEG, AmSO₄, and pH clear suites were performed as described in section 2.15.1, using the sample of BPSL2418 preparation 1, with 20 mg/ml concentration. The initial screens were accomplished then were incubated at 17 °C. Initial hits (microcrystals) were observed in five conditions: B7 in PACT (0.2 M NaCl, 0.1 M MES at pH 6 and 20% w/v PEG 6000), B8 in PACT (0.2 M NH₄Cl, 0.1 M MES at pH 6 and 20% w/v PEG 6000), B9 in PACT (0.2 M LiCl, 0.1 M MES at pH 6 and 20% w/v PEG 6000) and B10 in PACT (0.2 M MgCl₂, 0.1M MES at pH 6 and 20% w/v PEG 6000) (Figure 4.14). Optimization of these conditions was made to produce large crystals, using the sample of BPSL2418 preparation 2, with 30 mg/ml concentration. Attempts of optimization were carried out by altering the PEG 6000 concentration (10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and 30 % w/v) and pH (5.0, 6.0 and 7.0) for small crystals that were found in B7, B8, B9 and B10 in PACT. For the initial hit that appeared in B7 JCSG, the optimization was done by altering the PEG 4000 concentration (1, 2, 4, 6, 8, 10, 12, 14 and 16%) and pH (3.6, 4.6 and 5.6).



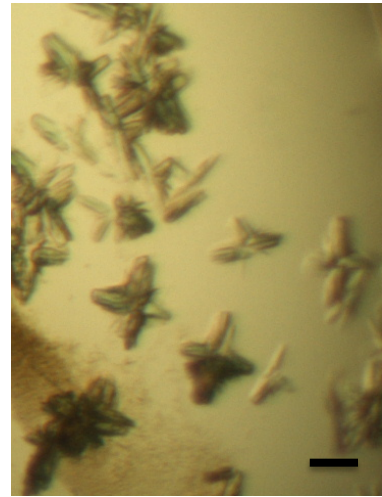
0.2 M NaCl, 0.1 M MES at pH 6 and
20% PEG 6000



0.2 M NH₄Cl, 0.1 M MES at pH 6 and
20% PEG 6000



0.2 M LiCl, 0.1 M MES at pH 6 and
20% PEG 6000



0.2 M LiCl, 0.1 M MES at pH 6 and
20% PEG 6000

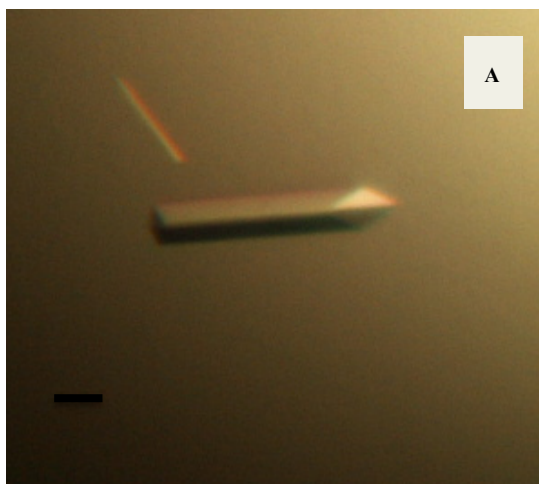
Figure 4.14: photographs of initial crystallization hits of BPSL2418_{MES} in five different conditions. Scale bar: 100 μ m.

The hanging-drop method (see sections 3.3.1.1) was utilized to optimize the initial hits and yield large crystals that were suitable for X-ray data collection. The droplet in the hanging-drop technique was made by combining 2 μ l protein and 2 μ l crystallization buffer and then hanging over a well of 1 ml of reservoir solution at 17 °C. Fortunately attempts to optimize crystallization conditions for BPSL2418 produced crystals in many different conditions. The best crystals of BPSL2418_{MES} were obtained with a rod shape in three concentrations of PEG 6000 conditions (18-20%), 0.2 M NaCl at pH 6 and 0.1 M MES (Figure 4.15). Desirable, rod shaped crystals were also seen in these two conditions (0.2 M NH₄Cl, 0.1 M MES at pH 6 and 20% w/v PEG 6000) and (0.2 M LiCl, 0.1 M MES at pH 6 and 18% w/v PEG 6000) (Figure 4.15).

4.6 Crystallization of BPSL2418 unbound, reduced and substrate-bound forms

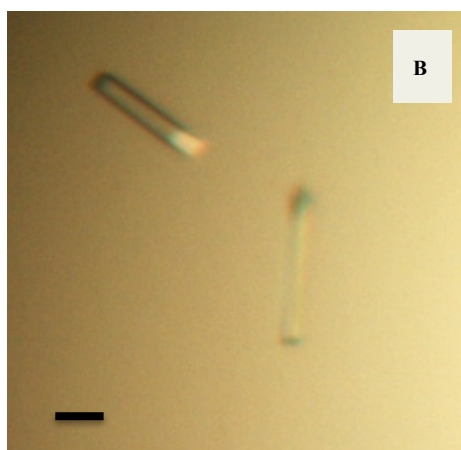
4.6.1 Crystallization of BPSL2418 unbound form (BPSL2418_{unbound})

It has been predicted from the sequence analysis that BPSL2418 may act as a free methionine-*R*-sulfoxide reductase. Therefore the crystallization buffers should contain no sulfoxide or sulfonic acid in order to produce an unbound form of BPSL2418. As mentioned above, most of the successful trials contained MES buffer and thus the only suitable condition that could produce unbound BPSL2418 is B7 in the JCSG screen (0.1 M sodium acetate at pH 4.6 and 8% PEG 4000) (Figure 4.16). Optimization of this condition was run in order to obtain large crystals. Unfortunately, no crystals were grown in the optimization screens.



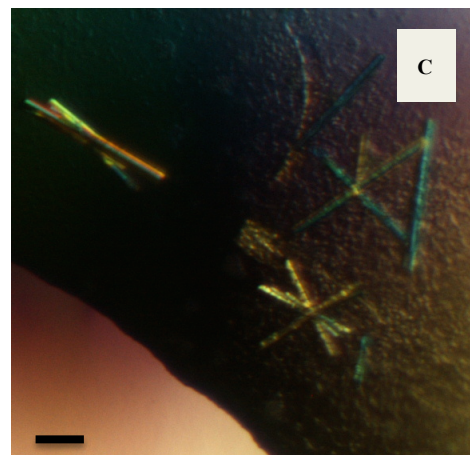
0.2 M NaCl, 0.1 M MES at pH 6 and
20% PEG 6000

Drop size 2 μ l protein + 2 μ l well



0.2 M NH₄Cl, 0.1 M MES at pH 6 and
20% PEG 6000

Drop size 2 μ l protein + 2 μ l well



0.2 M LiCl, 0.1 M MES at pH 6 and
20% PEG 6000

Drop size 2 μ l protein + 2 μ l well

Figure 4.15: Photographs of BPSL2418_{MES} crystals. (A) and (B) Most crystals had a rod shape; (C) cluster of rod shaped crystals was seen in a few wells. Scale bar: 100 μ m.

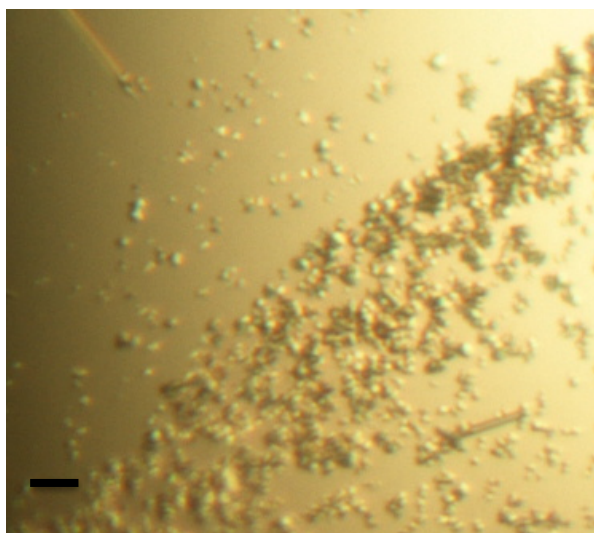


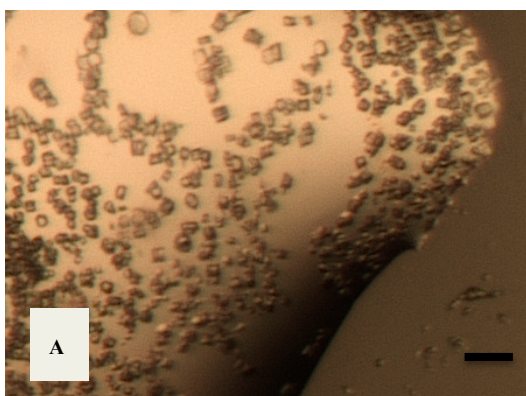
Figure 4.16: Small crystals of BPSL2418_{unbound} form produced from 0.1 M sodium acetate at pH 4.6 and 8% PEG 4000. Scale bar: 100 μ m.

4.6.2 Crystallization of BPSL2418 substrate-bound form (BPSL2418_{Met-SO})

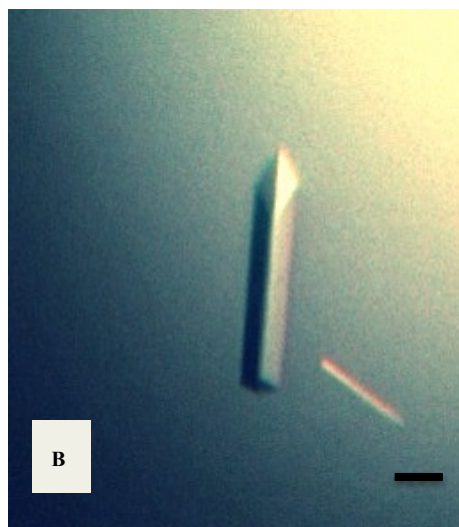
The structure analyses of BPSL2418 homologous proteins suggest that conformational changes occur in the active site (see Figure 1.7 and 1.8), where free methionine sulfoxide (Met-*R*-SO) is proposed to bind. In order to understand the role of BPSL2418, attempts to crystallize the BPSL2418 in complex with the free Met-*R*-SO substrate were performed. A mixture containing 20 mg/ml pure BPSL2418, 9 mM, free Met-*R*-SO was set up and incubated for 3 h at room temperature. Initial trials were carried out using this mixture, and hits were seen in 0.2 M ammonium chloride 0.1 M sodium acetate pH 5 and 20% PEG 6000 (Figure 4.17a). Crystals of the complex with free Met-*R*-SO were obtained from 0.2 M ammonium chloride, 0.1 M sodium acetate pH 6 and 18% PEG 6000 (Figure 4.17b). One crystal was soaked in cryoprotectant solution containing 30% ethyleneglycol and 0.2 M ammonium chloride, 0.1 M sodium acetate pH 6 and 18% PEG 6000. The crystal diffracted in-house to better than 2.5 Å resolution, and then it was frozen in liquid nitrogen to be sent to the Diamond Light Source.

4.6.3 Crystallization of BPSL2418 reduced form (BPSL2418_{Reduced})

Obtaining the structure of the reduced form of BPSL2418 can be valuable in understanding the role of BPSL2418. 10mM DTT (reducing agent) was added to a pure 20mg/ml BPSL2418 and the mixture was used to set up initial crystallization trials. Small crystals were found in 0.1M MIB buffer (this buffer is produced by mixing sodium malonate, imidazole and boric acid in the molar ratios 2:3:3-sodium malonate:imidazole:boric acid) pH 4, 25% PEG 1500 (Figure 4.18a), and then optimization trials were performed as previously described to achieve large crystals (Figure 4.18b). Crystals with a reasonable size were selected and soaked in a cryoprotectant of 30% ethyleneglycol and 0.1 M MIB buffer pH 4, 23% PEG 1500. Crystal diffraction was initially tested in-house, where the best crystal diffracted to 3.1 Å. Best crystals were stored in liquid nitrogen and sent to the Diamond Light Source.

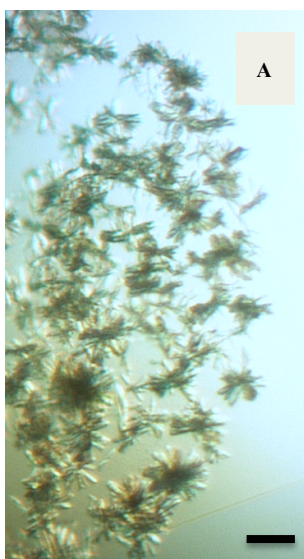


0.2 M ammonium chloride, 0.1 M sodium acetate pH 5 and 20% PEG 6000

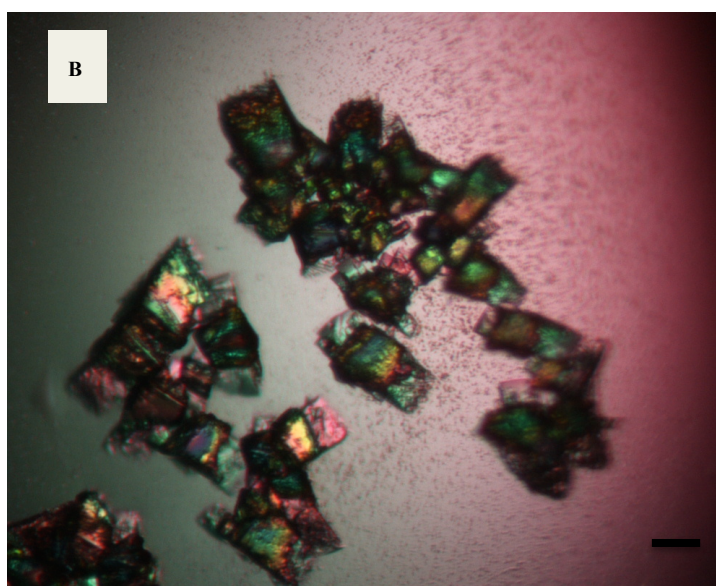


0.2 M ammonium chloride, 0.1 M sodium acetate pH 6 and 18% PEG 6000

Figure 4.17: Photographs of BPSL2418_{Met-SO} complex crystals. A: Initial hits; B: Crystals. Scale bar: 100 μ m.



0.1 M MIB buffer pH 4,
25% PEG 1500



0.1 M MIB buffer pH 4, 23% PEG 1500

Figure 4.18: Photograph of crystals from BPSL2418_{Reduced} form. A: Initial small crystals. B: clusters of crystals. Scale bar: 100 μ m.

4.7 Cryo-protection for BPSL2418 forms

Cryoprotection as mentioned in section 3.4.2 is a very significant step that helps to reduce radiation damage. The exposure of the crystal to the X-ray radiation can produce free radicals, which distribute inside the crystal and damage the protein. Therefore, a cryoprotectant solution should be selected very carefully and different conditions need to be tested as described in section 2.15.2 to choose the optimum cryoprotectant condition. It was found that the best cryoprotectant solution for the harvested crystals of the BPSL2418 forms is 30% ethylene glycol mixing with precipitant. The initial data collection for the BPSL2418 crystals were carried out on the in-house source utilizing the Rigaku Micromax 007 copper rotating anode generator with a Mar345 detector. In order to achieve higher resolution, the crystals were stored in liquid nitrogen at 100 k to send it to the Diamond Light Source in Oxford for data collection.

Chapter 5: Structure determination of BPSL2418

In this chapter a description of the structure solutions of BPSL2418_{MES}, BPSL2418_{Met-SO} and BPSL2418_{Reduced} by molecular replacement will be provided. BPSL2418 shows a very high sequence similarity, 68% to the *Staphylococcus aureus* fRMs_r, which allowed the use of this protein as a model in initial phasing by molecular replacement.

5.1 Experimental structure determination of BPSL2418_{MES} by molecular replacement using PDB entry: 3ksf.

5.1.1 Space group determination and indexing

Initial diffraction analysis of BPSL2418_{MES} crystals was carried out in Sheffield using the Rigaku Micromax 007 copper rotating anode generator with MAR 345 image plate, in order to evaluate the diffraction quality. 1.542 Å wavelength X-rays were used and the crystal diffracted to 2.0 Å. Two images were collected of 1° oscillations separated by 90° with 20 minutes exposure each. These initial test images were used to find out the lattice type by auto-indexing using MOSFLM (Leslie, 1994). This was confirmed by collecting three test images 45° apart with 1° oscillation at the Diamond Light Source in Oxford using the auto-indexing and collection strategy components of MOSFLM. BPSL2418_{MES} belongs to the primitive orthorhombic crystal system, which is distinguished by unequal length of the three axes a, b and c and the three interaxial angles equal to 90°, ($a \neq b \neq c$, $\alpha = \beta = \gamma = 90^\circ$). The Laue group is P 2 2 2 and the crystal had a cell dimensions $a = 53.48$ Å, $b = 60.54$ Å, $c = 42.24$ Å, $\alpha = 90^\circ$, $\beta = 90^\circ$ and $\gamma = 90^\circ$.

5.1.2 Data collection and processing

The BPSL2418_{MES} crystals, were saved in liquid nitrogen and data were collected at 100 k on the I04 beamline of the Diamond Light Source, Oxford. Three test images 45° apart with 1° oscillation were collected as described in section 5.1.1 to confirm crystal centering and to obtain a collection strategy using MOSFLM (Leslie, 1994). A total of 360 images were collected with 0.5° rotation per image at a crystal-to-detector distance of 250 mm at a wavelength 0.9795 Å. The exposure time was 0.5 second per image at 40% beam-transmission. Data extending to 1.18 Å were collected (Figure 5.1), and the data were processed using xia2 (Winter, 2010), indexed and integrated by XDS (Kabsch, 2010a) and scaled by XSCALE (Kabsch, 1988). The processing data collection statistics of BPSL2418_{MES} at 1.18 Å resolution are shown in table 5.1. The result of data collection and processing gave the overall R_{merge} 0.072 (0.566 for the outer shell), the $I/\sigma I$ was overall 16.3 (2.6 for the outer shell) and the completeness was overall 97.3%. Also the data do not appear to be twinned and there was no sign of significant radiation damage.

5.1.3 Matthews Coefficient (V_m) calculation

Matthews Coefficient calculations allow an estimation of the number of molecules in the asymmetric unit (Matthews, 1968). The Matthews Coefficient was estimated using Mattprob (<http://www.ruppweb.org/Mattprob/>) (Kantardjieff & Rupp, 2003). This calculated value depends on the unit cell volume and the molecular weight of BPSL2418_{MES}. Mattprob showed one possible solution of one protein molecule in the asymmetric unit with a solvent content of 35.2% and a V_m of 1.9 Å³/Da (Figure 5.2).

5.1.4 Structure determination of BPSL2418_{MES}

As described in section 1.2.4 BPSL2418_{MES} shows high homology with the GAF domain fRMsr enzymes, which made phase determination by molecular replacement with a search model of one of these enzymes the first option.

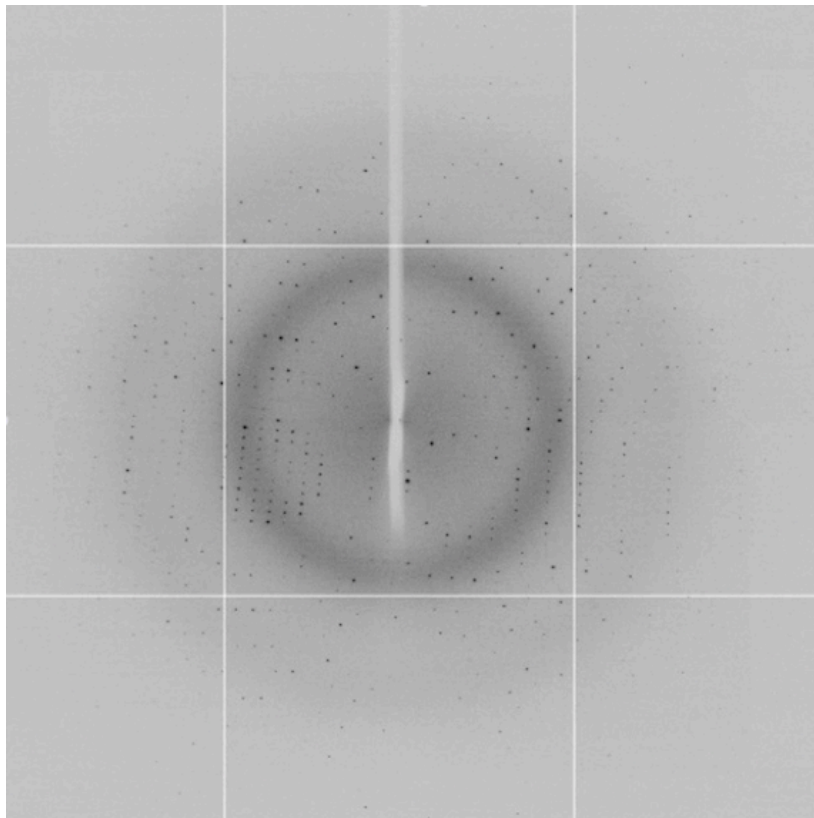


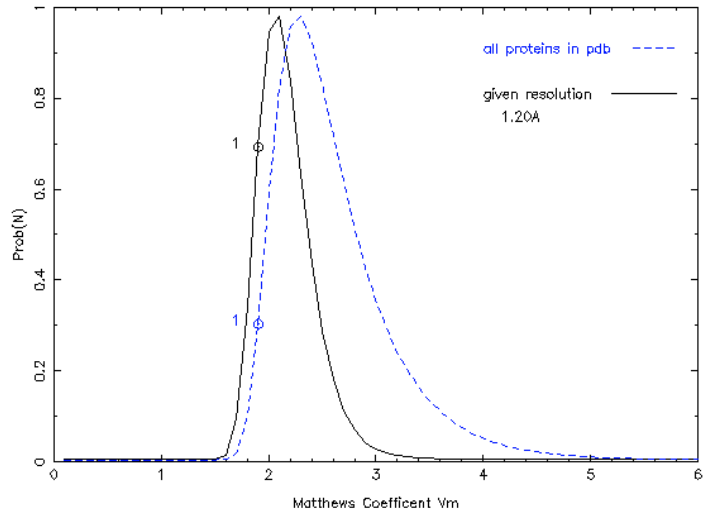
Figure 5.1: X-ray diffraction image of BPSL2418_{MES} crystal to 1.18 Å resolution using the I04 beamline of the Diamond Light Source.

Data collection statistics	
Beamline at DLS,Oxford, UK	I04
Detector	ADSC Q315r detector
The crystal-to-detector distance	250 mm
Space group	P 2 ₁ 2 ₁ 2 ₁
Crystal system	orthorhombic
The unit cell parameters:	
a (Å)	42.240
b (Å)	53.480
c (Å)	60.540
α (°)	90.000
β (°)	90.000
γ (°)	90.000
Wavelength (Å)	0.97950
Resolution range (Å)	60.5-1.18 (1.21-1.18)
No. of unique reflections	40778 (1566)
R_{merge}	0.07 (0.57)
R_{pim}	0.03 (0.29)
Mean (I)/σ(I)	16.3 (2.6)
Completeness (%)	97.3 (89.3)
Multiplicity	7.0 (5.4)
Mosaicity (°)	0.1
Number of molecules in the asymmetric unit	1

Table 5.1: Data collection statistics for the BPSL2418_{MES} number in parentheses shows the highest resolution shell.

A	Found 1 possible solutions :					
	Vm of protein (Matthews coeff.) is 1.90 A**3/Dalton					
	Estimated solvent content is 35.24 %					
	N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
	1	1.0000	1.0000	1.90	35.24	18000.00

B



C

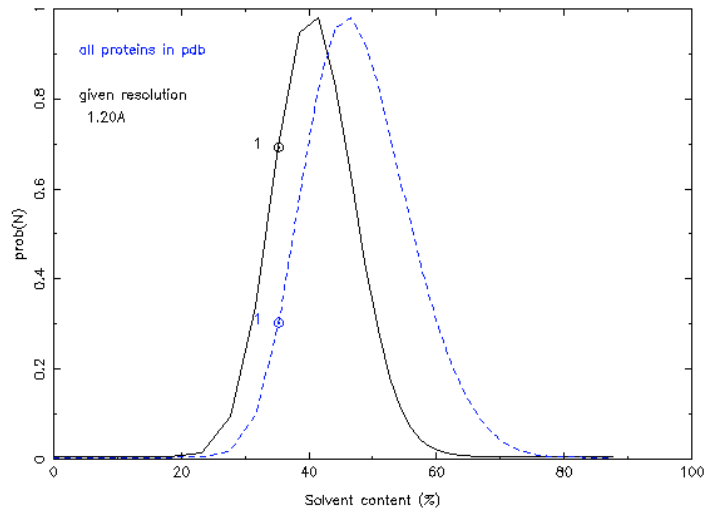


Figure 5.2: Matthews coefficient calculations and probabilities for BPSL2418_{MES} using Mattprob <http://www.ruppweb.org/Mattprob/>. (A) This indicates the possibility of one protein molecule in the asymmetric unit with a V_m of 1.9 Å/Da (B) and a solvent content of 35.2% (C).

The Phyre2 server (Kelley & Sternberg, 2009) was used to obtain a search model. A *Staphylococcus aureus* fRMs_r structure (PDB 3KSF) with 49% identity and 68% similarity was selected and used for an automated search in Phaser (McCoy *et al*, 2007). No modification was done before using the template structure. The 1.18 Å dataset (the scaled mtz file) was input and Phaser was run in P 2 2 2 with the all alternative space groups, searching for one molecule in the asymmetric unit. A single solution was obtained in space group P 2₁ 2₁ 2₁ with one molecule in the asymmetric unit, which correlated to the single model predicted by the Matthews probability calculations. This molecular replacement solution produced the rotation and translation function Z-scores of 8.1 and 13.1, respectively, which suggested the model was correct.

5.1.5 Model building and refinement

The model of BPSL2418_{MES} obtained by molecular replacement was viewed in COOT (Emsley & Cowtan, 2004) and immediately it was clear that the majority of the BPSL2418_{MES} sequence fitted well to the electron density. The molecular replacement solution produced R-factor and R_{free} of 0.409 and 0.428, respectively and after one round of refinement using REFMAC5 (Murshudov *et al*, 1997) the R-factor and R_{free} showed an improvement to 0.352 and 0.402, respectively. The refinement statistics exhibit a reduction in R-factor and R_{free} over 29 rounds of refinement and the model was rebuilt, refined and assessed using REFMAC5 and COOT (Figure 5.3). The refinement was carried on until there was no improvement could be seen. The final model composed of 162 residues, 180 water molecules and one MES molecule. The final refinement statistics for BPSL2418_{MES} structure are shown in table 5.2.

5.1.6 3D structure of BPSL2418_{MES}

The crystal structure of BPSL2418_{MES} shows that the protein has one molecule in the asymmetric unit as predicted by the Matthews coefficient calculations. This molecule was named chain A and the electron density of this molecule is well determined except for the first five residues at the N-terminus and the last three

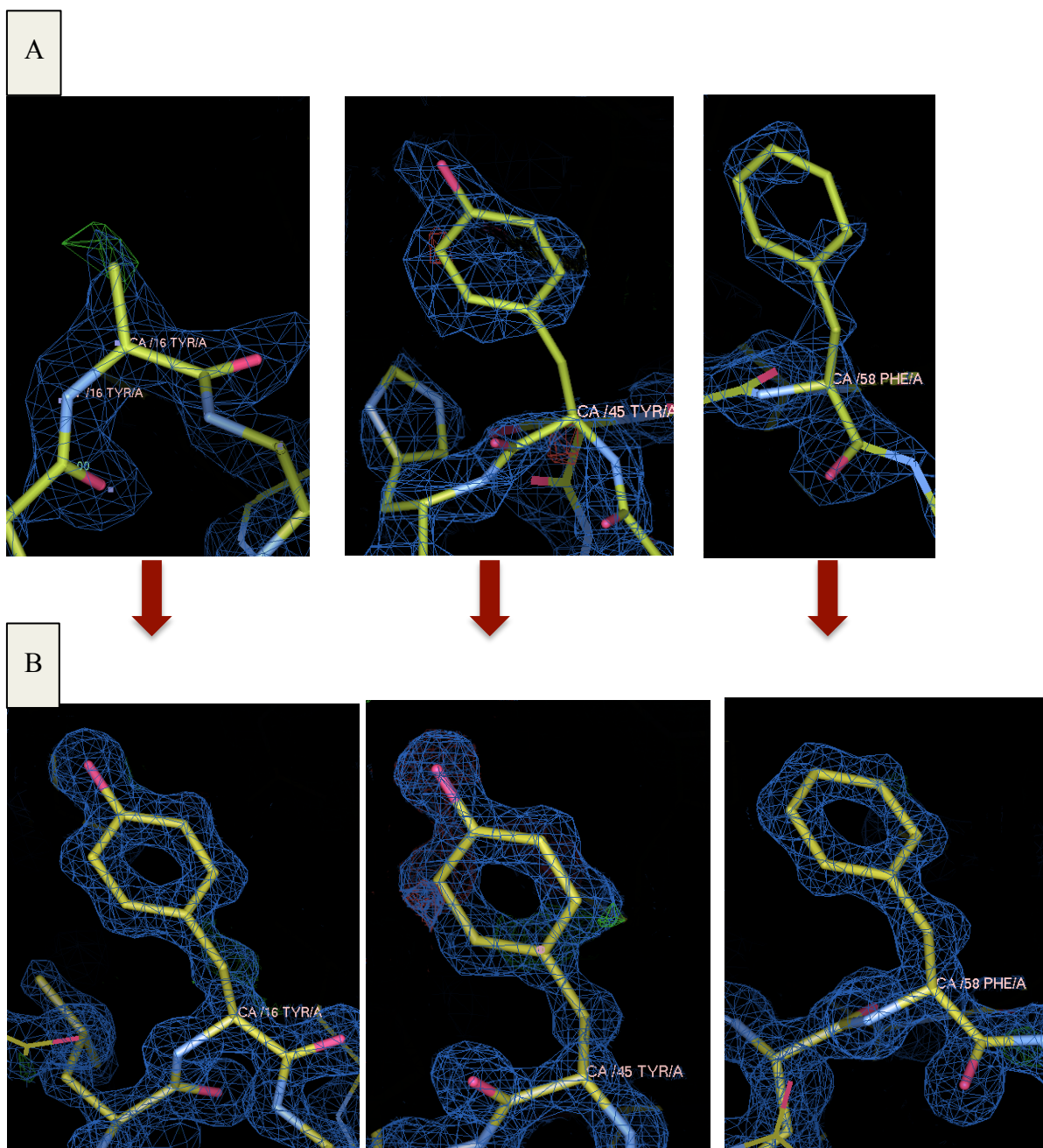


Figure 5.3: Sample regions of electron density for BPSL2418_{MES} at 1.18 Å show the structure improvement over 29 refinement cycles. (A) Shows three regions of electron density around residues 16,45 and 58 after one refinement. (B) Shows the same three regions after 29 cycles of rebuilding and refinement. Blue map calculated using coefficients $2mF_o-dF_c$ and contoured at 0.5σ .

Refinement statistics	
Resolution	1.18 Å
Protein molecules in asymmetric unit	1
Number of residues	162
Number of waters	180
Number of atoms	1383
Average B values (Å²):	
Whole chain A	8.52
Main chain	7.64
Side chain	9.57
Water	28.35
MES molecule	6.78
R.m.s deviation:	
Bond lengths	0.0100 Å
Bond angles	1.463°
Ramachandran plot:	
Most favored regions	98.75%
MolProbity score:	1.23/93 rd percentile
R-factor	0.12
R_{free}	0.17

Table 5.2: Refinement statistics for BPSL2418_{MES} after 29 cycles of rebuilding in COOT and refinement in REFMAC5. The Ramachandran scores were produced using MolProbity.

residues at the C-terminus. The refined electron density map shows a MES molecule binding the BPSL2418 close to CYS75 and CYS109 where a disulfide bond is formed between these two cysteines (Figure 5.4). BPSL2418_{MES} structure description and analysis will be discussed in chapter 6. The final model was validated using PROCHECK (Laskowski, 1993), and MOLPROBITY program (<http://molprobity.biochem.duke.edu>) (Chen et al, 2010). This shows that all residues fell within allowed regions of Ramachandran plot (Figure 5.5) and all main chain and side chain parameters were better or within the expected range for the resolution of data (Figure 5.6). These results reveal that the overall structure is of very good quality.

The PISA webserver <http://pdbe.org/PISA> (Krissinel & Henrick, 2007) was used to investigate the assemblies and interface between monomers in BPSL2418_{MES}. PISA was run using a model containing the BPSL2418 protein chain binding MES molecule and symmetry-related molecules generated by the three 2-fold screw symmetry axes in space group $P 2_1 2_1 2_1$. The PISA analysis shows that BPSL2418_{MES} monomers interact to form dimers (Figure 5.7) (this discussed in more details in chapter 6). The BPSL2418_{MES} crystal belongs to space group $P 2_1 2_1 2_1$ and the unit cell therefore contains four identical monomers, which suggests that the unit cell consist of 4 asymmetric units (Figure 5.8).

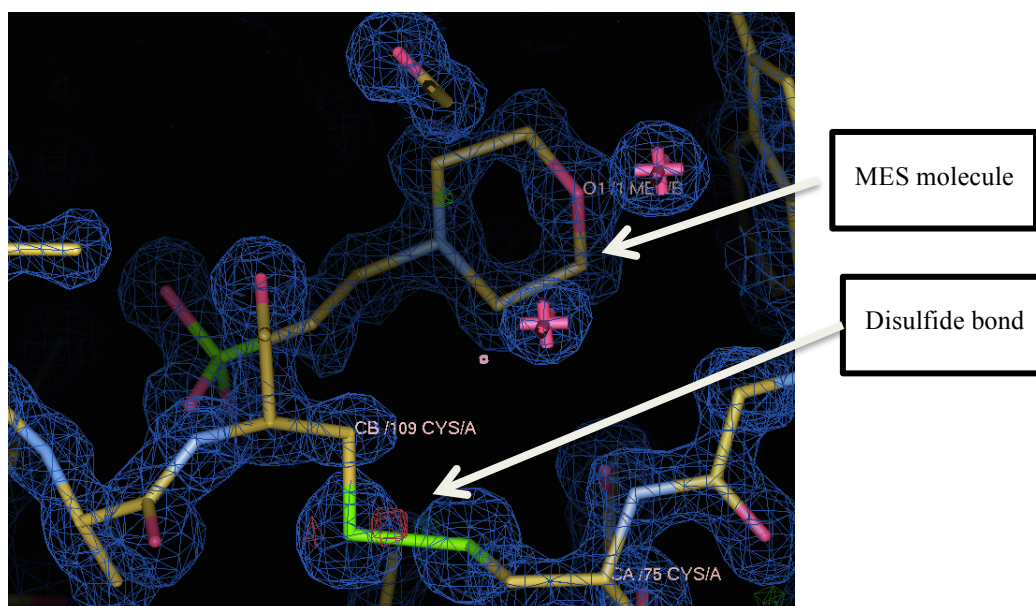
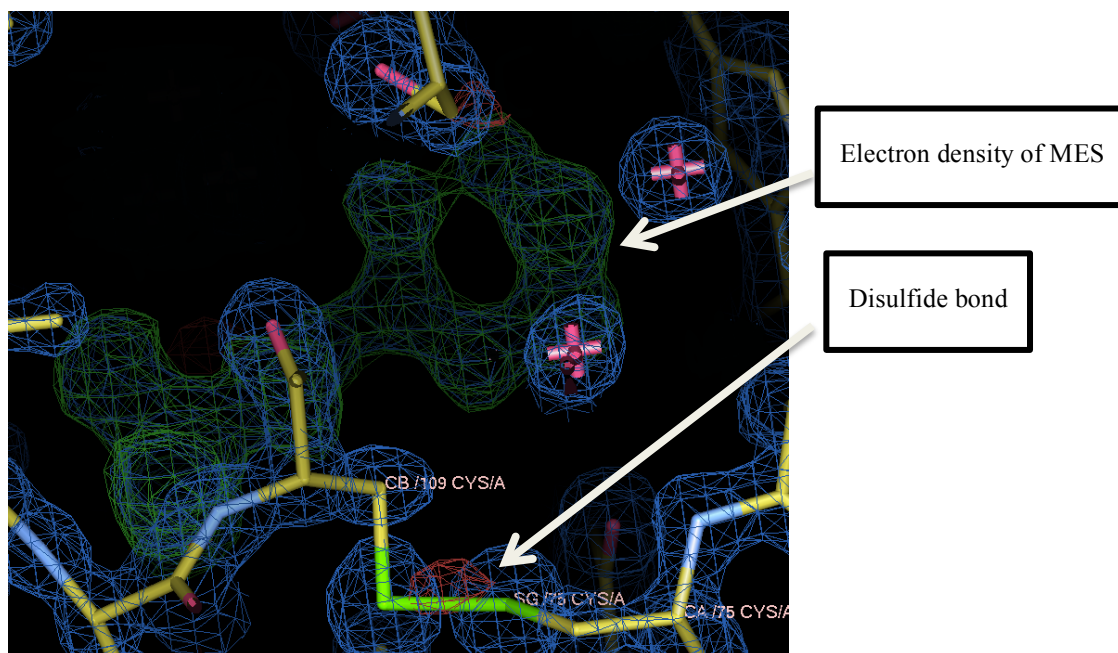
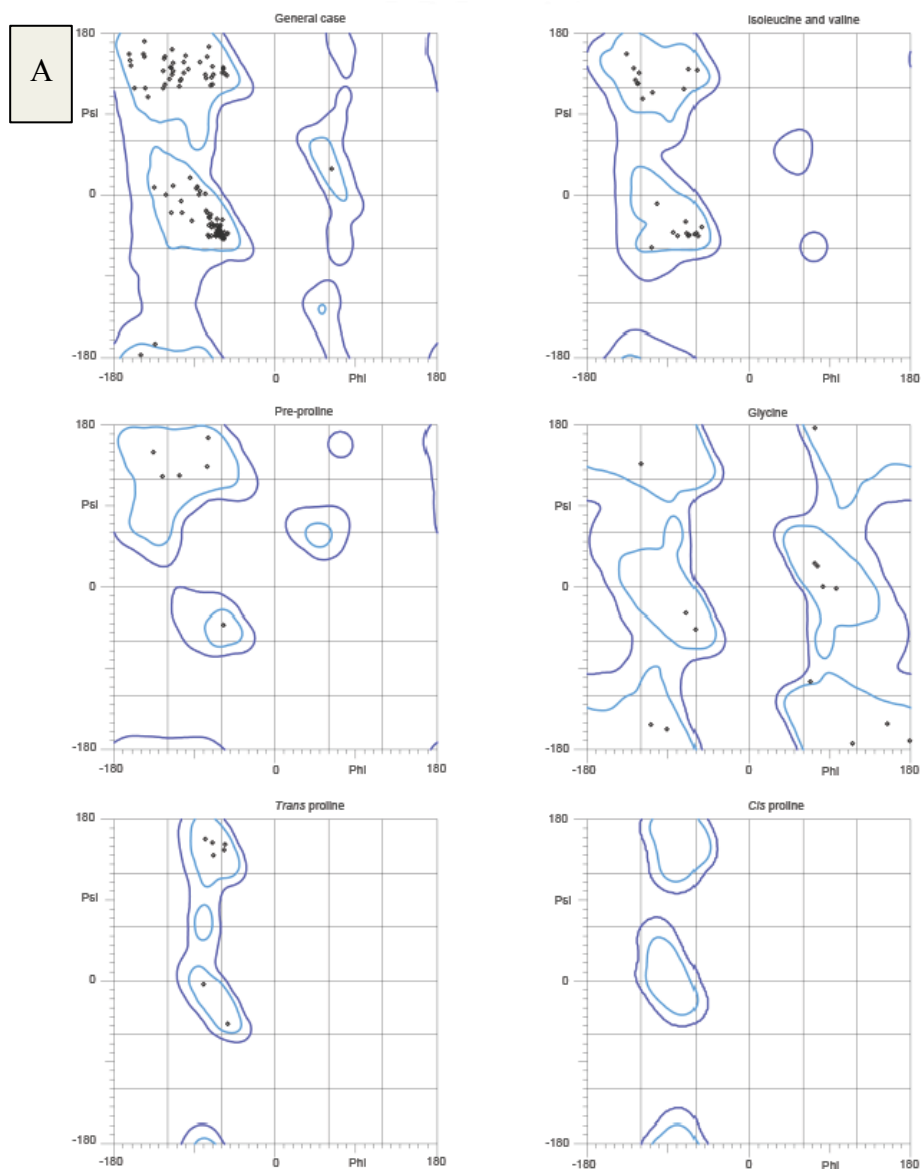


Figure 5.4: Electron density map of BPSL2418_{MES} at 1.18 Å shows a MES molecule binding the BPSL2418 close to CYS75 and CYS109 where a disulfide bond is formed between these two cysteines. Blue map calculated using coefficients $2mF_o-dF_c$ and contoured at 0.5σ . Green map calculated using coefficients mF_o-dF_c and contoured at 2.5σ .



98.8% (158/160) of all residues were in favored (98%) regions.
 100.0% (160/160) of all residues were in allowed (>99.8%) regions.

There were no outliers.

B

All-Atom Contacts	Clashscore, all atoms:	4.51	86 th percentile* (N=283, 1.20Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	0	0.00% Goal: <1%
	Ramachandran outliers	0	0.00% Goal: <0.05%
	Ramachandran favored	158	98.75% Goal: >98%
	MolProbity score [^]	1.23	93 rd percentile* (N=1439, 1.20Å ± 0.25Å)
	Cβ deviations >0.25Å	1	0.68% Goal: 0

Figure 5.5: Result of MOLPROBITY and Ramachandran plot of BPSL2418_{MES} structure. (A) The Ramachandran plot shows that all residues are within the favored region. (B) Overall the MOLPROBITY score is in the 93rd percentile (100% being the best amongst structures of comparable resolution). MOLPROBITY (Chen *et al*, 2010).

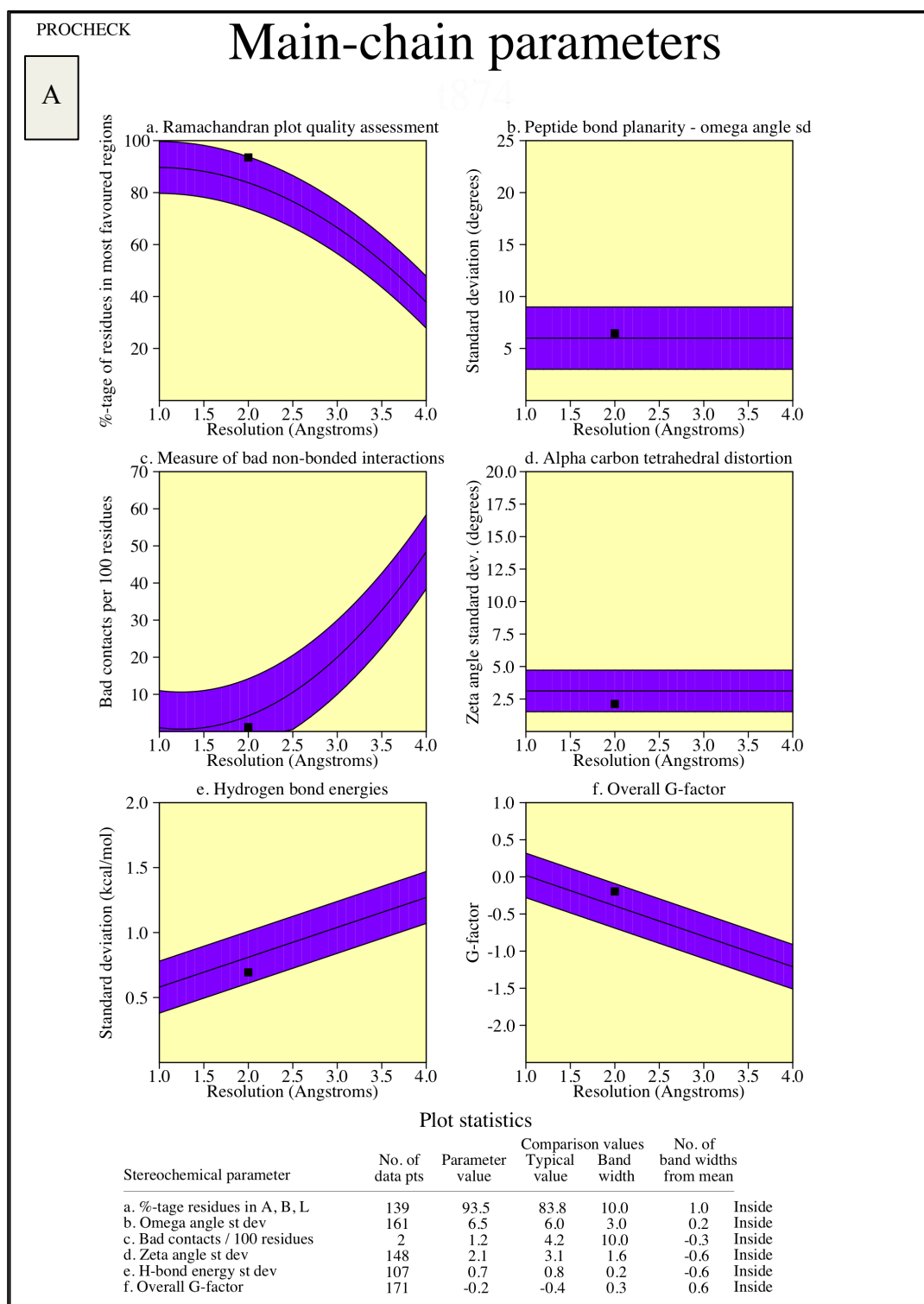
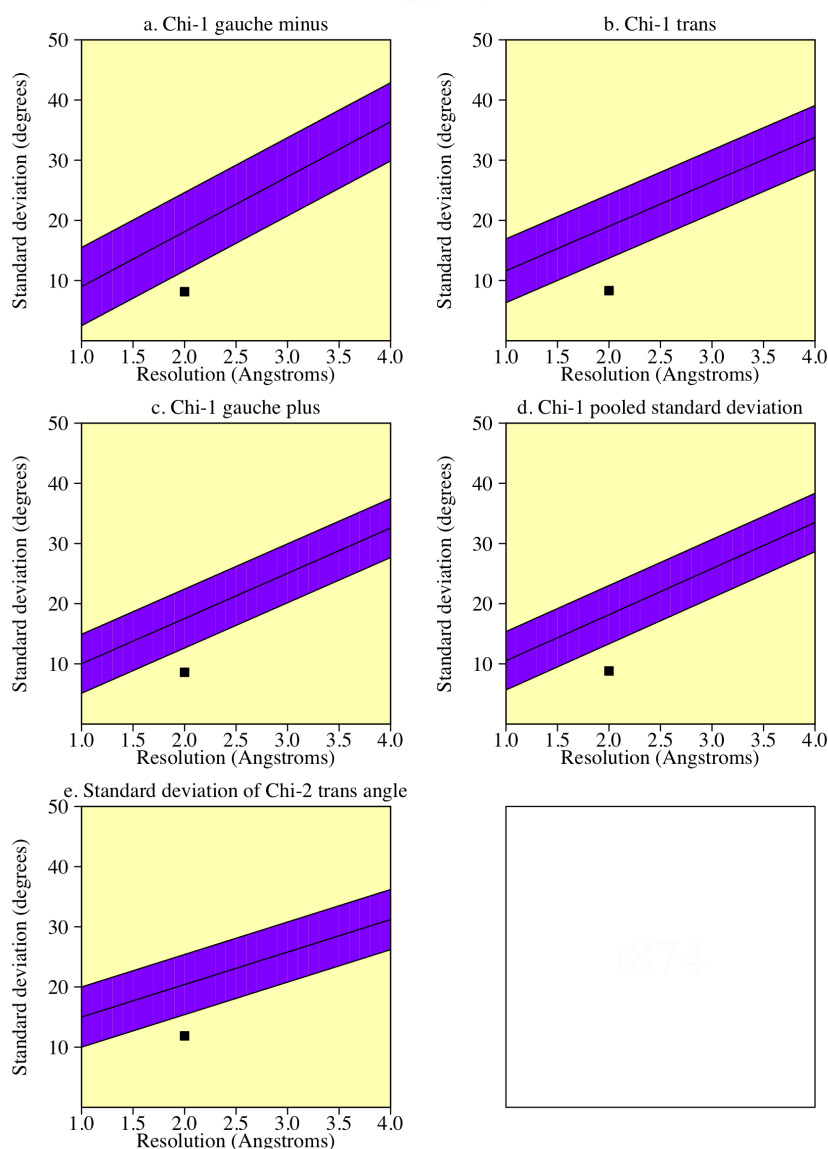


Figure 5.6: Main chain and side chain parameters of BPSL2418_{MES}. (A) Showing all the main chain parameters values are within the average for a structure at this resolution.

Side-chain parameters

B

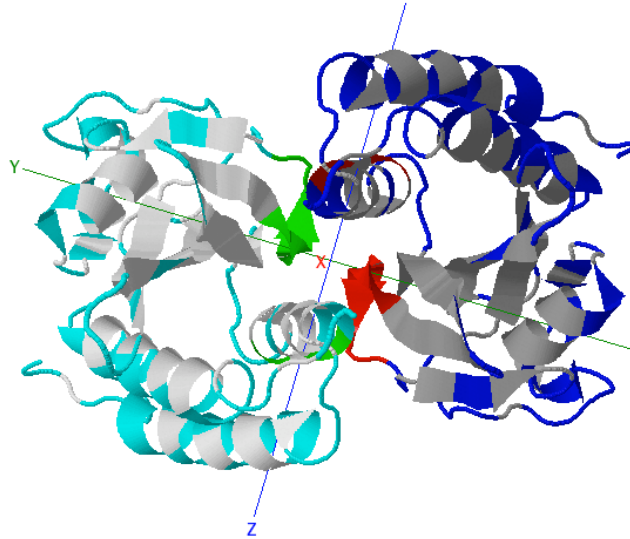


Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean	
			Typical value	Band width		
a. Chi-1 gauche minus st dev	17	8.1	18.1	6.5	-1.5	BETTER
b. Chi-1 trans st dev	43	8.3	19.0	5.3	-2.0	BETTER
c. Chi-1 gauche plus st dev	51	8.6	17.5	4.9	-1.8	BETTER
d. Chi-1 pooled st dev	111	8.8	18.2	4.8	-1.9	BETTER
e. Chi-2 trans st dev	28	11.9	20.4	5.0	-1.7	BETTER

(B) Showing all the side chain parameters values are better than the average for a structure at this resolution. Produced by the programme PROCHECK (Laskowski *et al*, 1993).

A



B

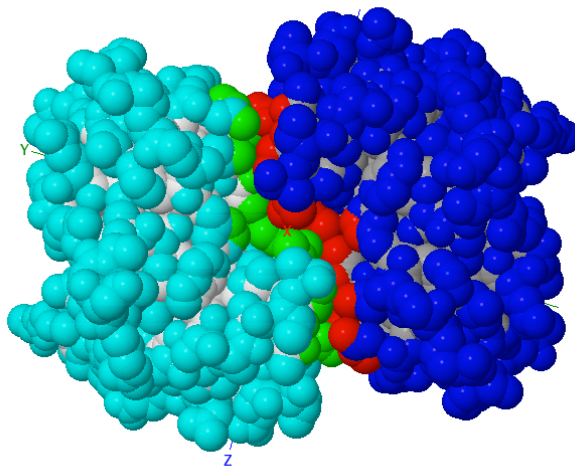
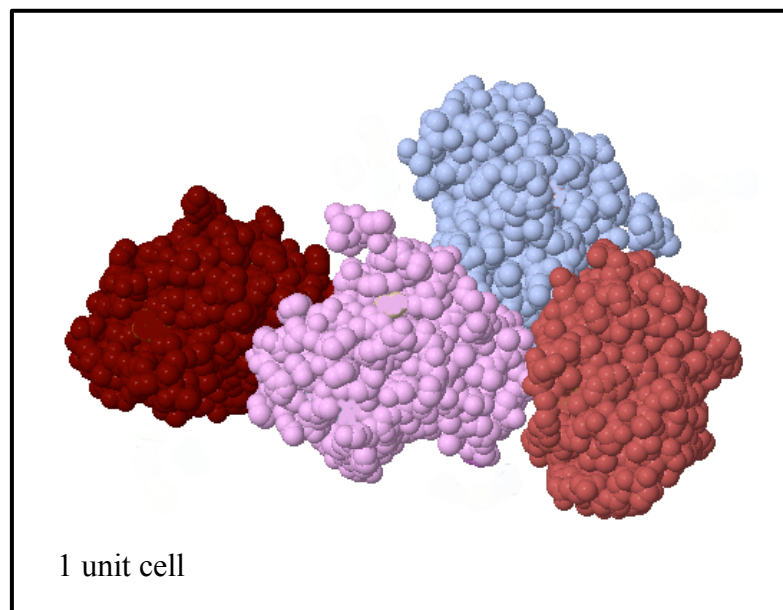


Figure 5.7: Dimeric BPSL2418_{MES}. (A) Cartoon model. (B) space-filling model. The Figure was produced using PISA webserver <http://pdbe.org/PISA>.

A



B

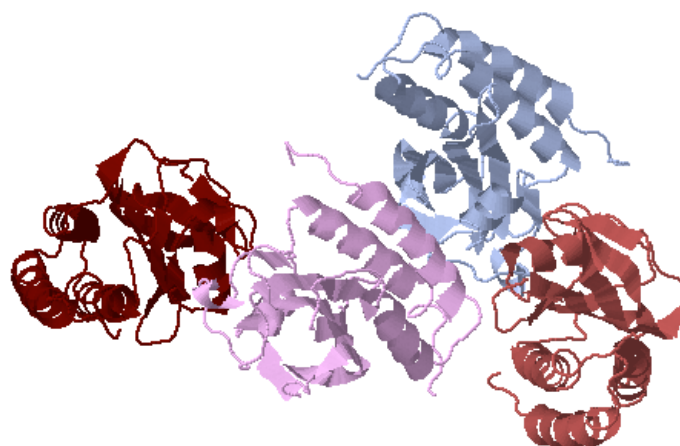


Figure 5.8: Four BPSL2418_{MES} molecules in a unit cell (space group $P 2_1 2_1 2_1$) of the crystal. (A) Spacefill model. (B) Cartoon model. The unit cell has two biological dimers and interaction between them is crystallographic interaction. The Figure was produced using PISA webserver http://pdbe.org/PISA_

5.2 Experimental structure determination of BPSL2418_{Met-SO} by molecular replacement using BPSL2418_{MES}

The best crystals of BPSL2418 in complex with the Met-SO substrate were selected from the optimization trial for data collection (section 4.6.2).

5.2.1 Space group determination and indexing

The initial data collection was carried out in Sheffield to estimate the diffraction quality. The best in-house diffraction of BPSL2418_{Met-SO} was 2.5 Å and to obtain better diffraction the crystal was sent to the Diamond Light Source at the I02 beamline. To obtain a collection strategy, three initial test images were collected 45° apart with 0.2° oscillation and analyzed using MOSFLM. BPSL2418_{Met-SO} belongs to the primitive monoclinic crystal system, which is distinguished by unequal length of the edges of the unit cell ($a \neq b \neq c$) and the equal α and γ angles of 90° and β unequal to 90° ($\alpha = \gamma = 90^\circ$, $\beta \neq 90^\circ$). BPSL2418_{Met-SO} data was indexed in P 1 2 1 space group with unit cell dimensions; $a = 35.90$ Å, $b = 92.27$ Å, $c = 42.65$ Å, $\alpha = 90^\circ$, $\beta = 102.42^\circ$ and $\gamma = 90^\circ$.

5.2.2 Data collection and processing

The BPSL2418_{Met-SO} data were collected at the I02 beamline of the Diamond Light Source, Oxford. 900 images were collected for the data set with 0.2° oscillation per image using X-ray of wavelength 0.97949 Å and using an Pitatus 6M detector with a crystal-to-detector distance of 246.4 mm. The crystal diffracted to 1.4 Å resolution (Figure 5.9). The collected images were processed using xia2 system and were indexed and integrated by XDS and scaled by XSCALE. All statistics of processed data collection of BPSL2418_{Met-SO} at 1.4 Å resolution are described in table 5.3. The overall R_{merge} was 0.079, (0.902 for the outer shell)

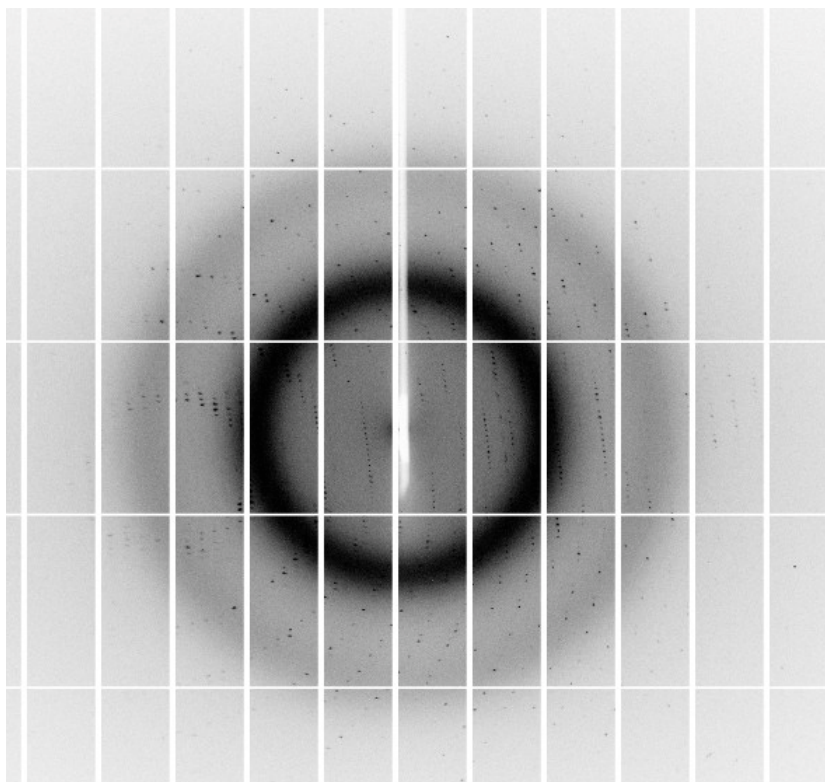


Figure 5.9: X-ray diffraction image of BPSL2418_{Met-SO} crystal to 1.4 Å using the I02 beamline of the Diamond Light Source.

Data collection statistics	
Beamline at DLS,Oxford, UK	I02
Detector	Pitatus 6M
The crystal-to-detector distance	246.40 mm
Space group	P2 ₁
Crystal system	monoclinic
The unit cell parameters:	
a (Å)	35.900
b (Å)	92.270
c (Å)	42.650
α (°)	90.000
β (°)	102.420
γ (°)	90.000
Wavelength (Å)	0.97949
Resolution range (Å)	32.8-1.42 (1.46-1.42)
No. of unique reflections	49084 (3603)
R_{merge}	0.08 (0.9)
R_{pim}	0.06 (0.63)
Mean (I)/σ(I)	9.7 (2.0)
Completeness (%)	96.2 (95.6)
Multiplicity	3.3 (3.4)
Mosaicity (°)	0.4
Number of molecules in the asymmetric unit	2

Table 5.3: Data collection statistics for the BPSL2418_{Met-SO}. Number in parentheses shows the highest resolution shell.

the $I/\sigma I$ was overall 9.7 (2 for the outer shell) and the completeness was overall 96.3%. Also the data do not appear to be twinned and there was no sign of radiation damage.

5.2.3 Matthews Coefficient (V_m) calculation

The Matthews Coefficient number V_m for BPSL2418_{Met-SO} was calculated using Mattprob webservice. The Matthews Coefficient calculation shows the asymmetric unit was estimated to include one or two protein molecules, with two being the most likely solution, producing a Matthews coefficient of 1.92 Å³/Da and a solvent content of 35.8% (Figure 5.10).

5.2.4 Structure determination of BPSL2418_{Met-SO}

The structure of BPSL2418_{MES} solved as part of this thesis was used as a search model for molecular replacement utilizing the diffraction data of BPSL2418_{Met-SO}. The 1.4Å dataset was input and Phaser was run in P2 with all alternative space groups, searching for two molecules in the asymmetric unit. A single solution was found in space group P2₁. This molecular replacement solution produced rotation and translation function Z-scores of 8.8 and 40.2 respectively, which suggested the model was correct.

5.2.5 Model building and refinement

The electron density map of BPSL2418_{Met-SO} produced by molecular replacement was viewed in COOT. The molecular replacement gave a model with an initial R-factor of 0.290 and R_{free} of 0.323. Reductions in R-factor and R_{free} to 0.18 and 0.20 respectively were achieved using 38 cycles of rebuilding and refinement and showed an improvement in the fit of the model to the map (Figures 5.11 & 5.12). The refinement was carried on until there was no possible structure improvement. The final model was composed of two molecules chain A and chain B, which consist of 155 residues and 156 residues, respectively. The final refinement statistics for the BPSL2418_{Met-SO} structure are shown in table 5.4.

Found 2 possible solutions :

A	Vm of protein (Matthews coeff.) is	3.83 $\text{\AA}^3/\text{Da}$
	Estimated solvent content is	67.91 %
	Vm of protein (Matthews coeff.) is	1.92 $\text{\AA}^3/\text{Da}$
	Estimated solvent content is	35.82 %

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm $\text{\AA}^3/\text{Da}$	Vs % solvent	Mw Da
1	0.0106	0.1696	3.83	67.91	18000.00
2	0.9894	0.8304	1.92	35.82	36000.00

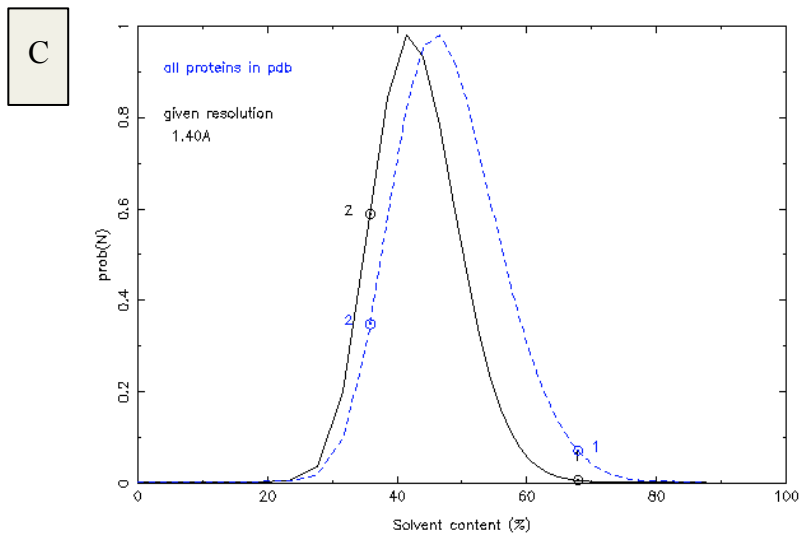
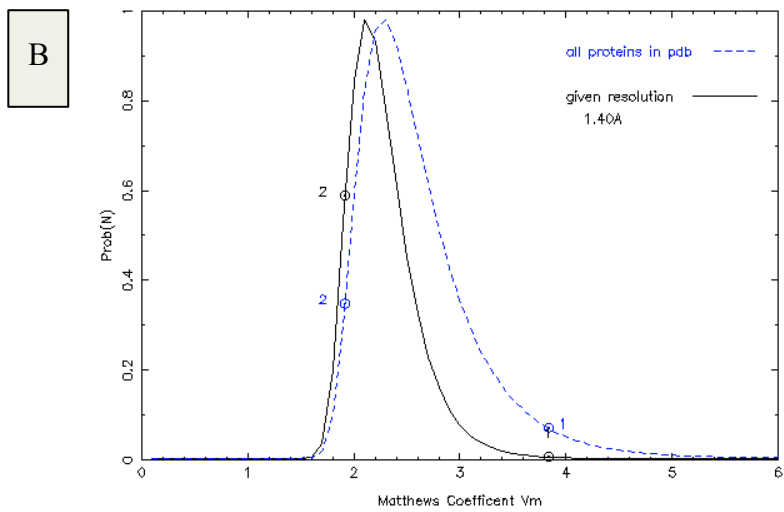


Figure 5.10: Matthews coefficient calculations and probabilities for BPSL2418_{Met-SO} using Mattprob <http://www.ruppweb.org/Mattprob/>. This indicates the possibility of two protein molecules in the asymmetric unit (A) with a Vm of 1.92 $\text{\AA}^3/\text{Da}$ (B) and a solvent content of 35.8% (C)

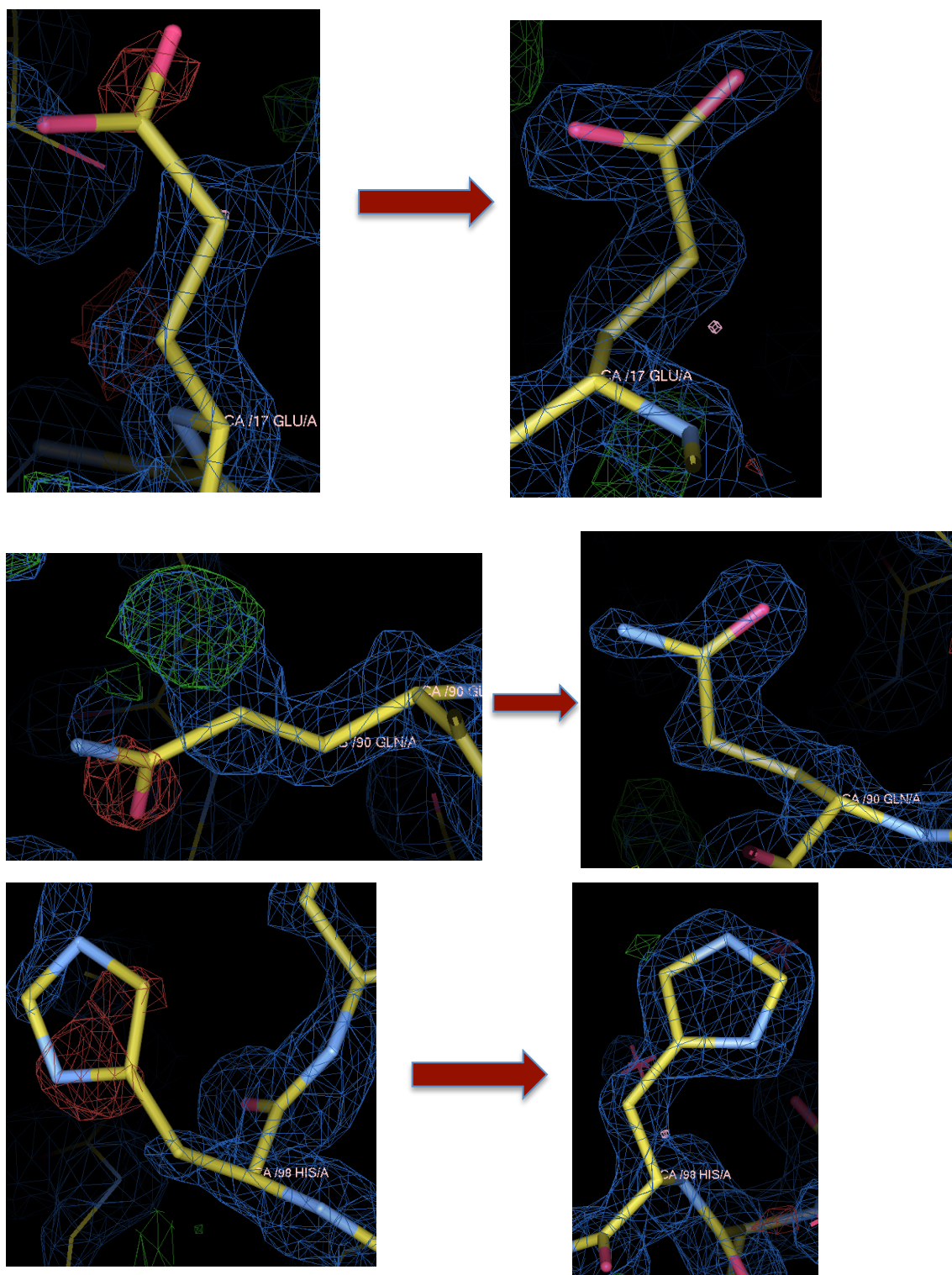


Figure 5.11: Sample regions of electron density for BPSL2418_{Met-SO} chain A at 1.4 Å show the structure improvement over 38 refinement cycles. On the left side three regions of electron density around residues 17,90 and 98 after one cycle of refinement. On the right side the same three regions after 38 cycles of rebuilding in COOT and refinement in REFMAC5. Blue map calculated using coefficients 2mFo-dFc and contoured at 0.4 σ . Green map calculated using coefficients mFo-dFc and contoured at 2.4 σ .

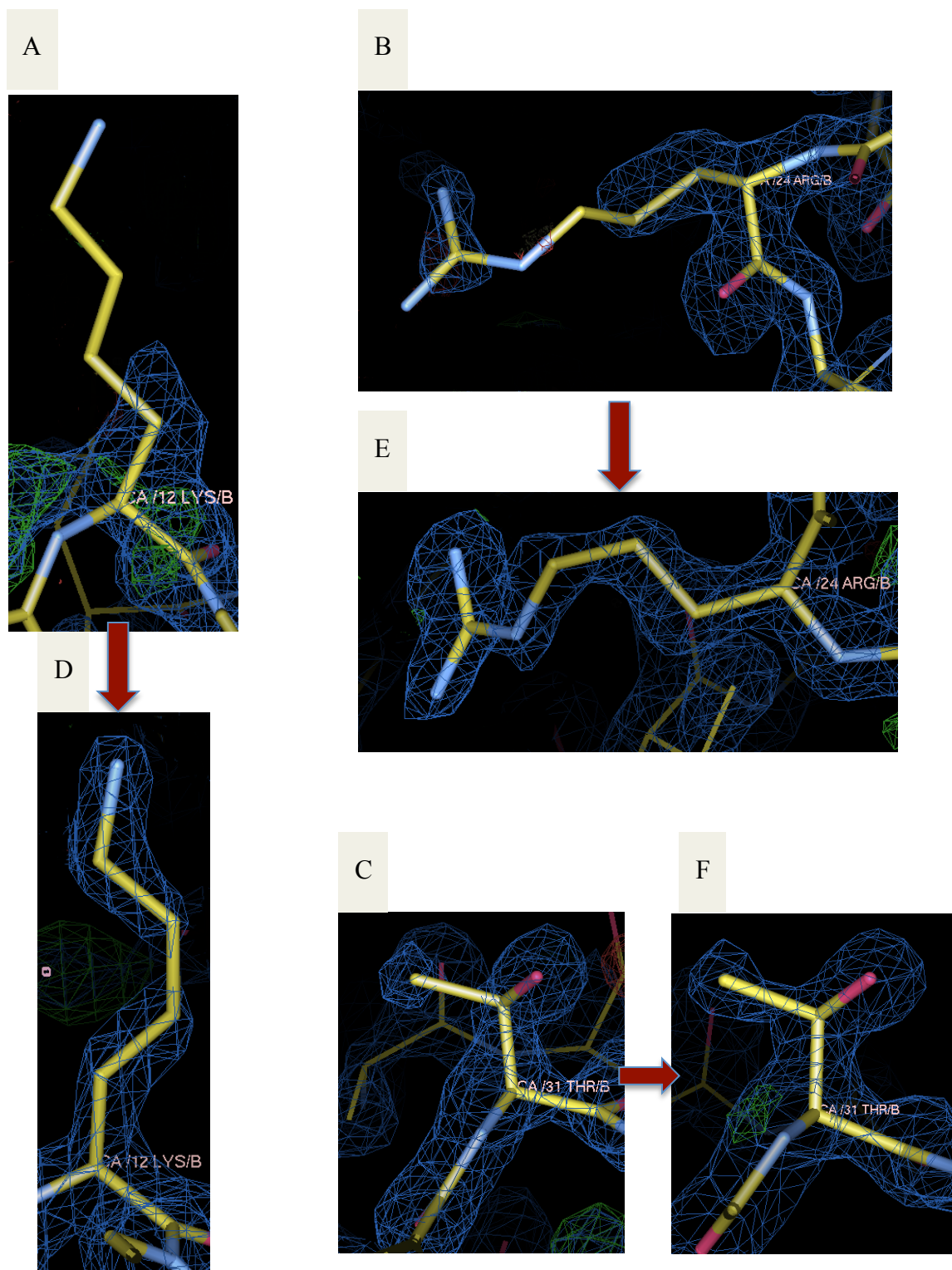


Figure 5.12: Sample regions of electron density for BPSL2418_{Met-SO} chain B at 1.4 Å show the structure improvement over 38 refinement cycles. (A),(B) and (C) Three regions of electron density around residues 12,24 and 31 after one cycle refinement. (D),(E) and (F) The same three regions after 38 cycles of rebuilding in COOT and refinement in REFMAC5. Blue map calculated using coefficients 2mFo-dFc and contoured at 0.4 σ . Green map calculated using coefficients mFo-dFc and contoured at 2.4 σ

5.2.6 3D structure of BPSL2418_{Met-SO}

The crystal structure of BPSL2418_{Met-SO} shows that the protein has two molecules in the asymmetric unit as predicted by the Matthews coefficient calculations. The electron density of these molecules is well determined except for ten residues in the N-terminus for both molecules and the last four or five residues in the C-terminus for chain A and chain B, respectively. The electron density map shows a Met-SO molecule binding the BPSL2418 in the active site close to CYS75 and CYS109 in both chain A and chain B. A disulfide bond is formed between these two cysteines in both molecules A and B (Figure 5.13). BPSL2418_{Met-SO} structure description and analysis will be discussed in chapter 6. The final model was validated using PROCHECK (Laskowski, 1993), and MOLPROBITY program (Chen *et al*, 2010). This shows that all residues fell within allowed regions of Ramachandran plot (Figure 5.14) and all main chain and side chain parameters were better or within the expected range for the resolution of data (Figure 5.15). These results reveal that the overall structure is of very good quality.

To identify the assemblies and interface between monomers in BPSL2418_{Met-SO} the PISA webserver was used. PISA was run using a model involving the BPSL2418 chain binding Met-SO substrate and it indicates the presence of a dimer interface between the two monomers of BPSL2418_{Met-SO} A and B (this is discussed in more chapter 6) (Figure 5.16). The BPSL2418_{Met-SO} crystal belongs to space group P2₁ which has one 2-fold screw axis and thus the unit cell contains two dimers of BPSL2418_{Met-SO} (Figure 5.17).

Refinement statistics	
Resolution	1.4 Å
Protein molecules in asymmetric unit	2
Number of atoms	2542
Number of residues	311
Number of waters	390
Average B values (Å²):	
Whole chain A	14.6
Main chain A	13.5
Side chain A	15.9
Met-SO molecule	12.9
Whole chain B	15.6
Main chain B	14.6
Side chain B	16.8
Met-SO molecule	13.2
Water	30.9
R.m.s deviation:	
Bond lengths	0.0136 Å
Bond angles	1.631°
Ramachandran plot:	
Most favored regions	98.75%
MolProbity score:	1.44/78 th percentile
R-factor	0.17
R_{free}	0.19

Table 5.4: Refinement statistics for BPSL2418_{Met-SO} after 38 cycles of rebuilding in COOT and refinement in REFMAC5. The Ramachandran scores were produced using MolProbity.

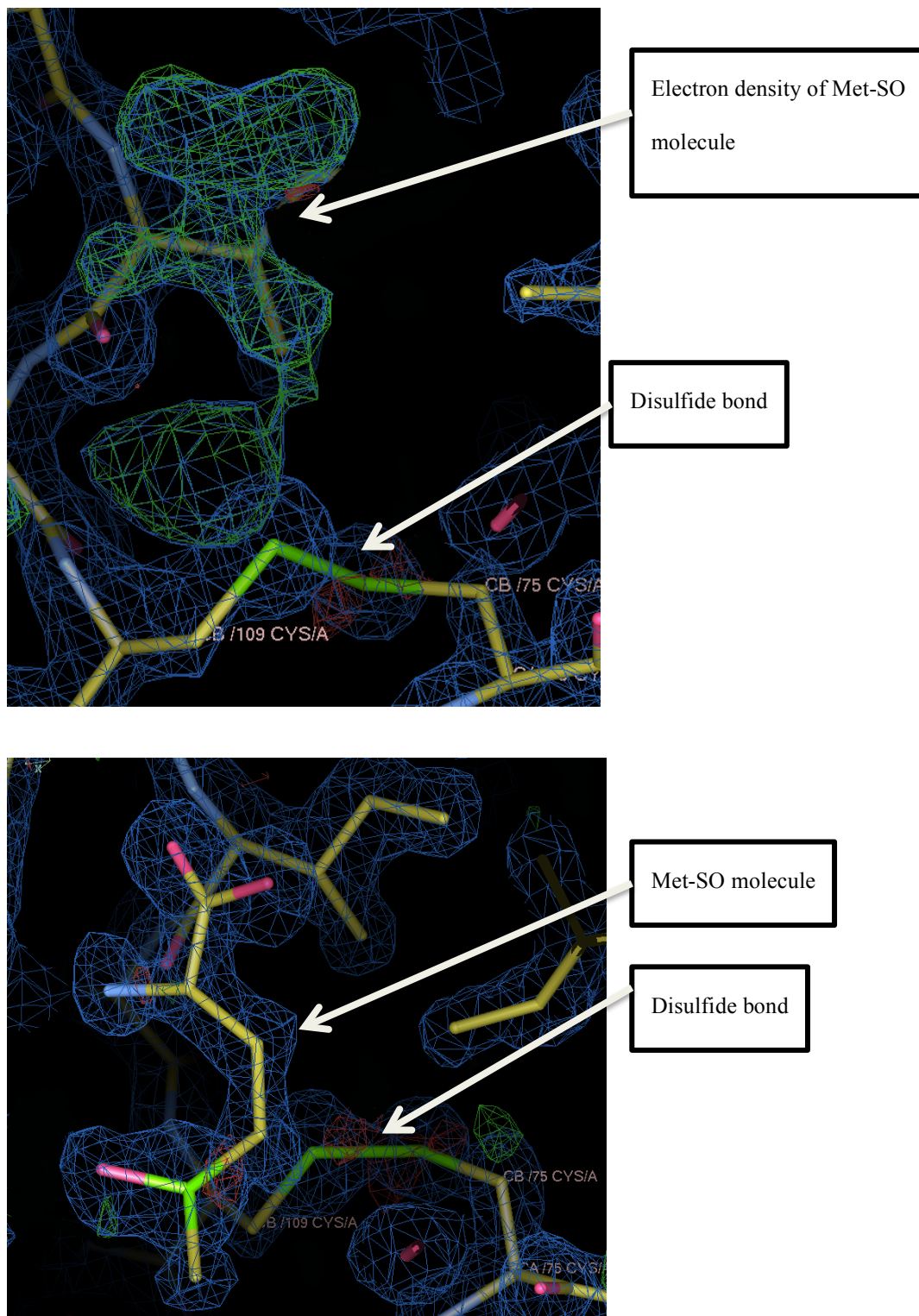
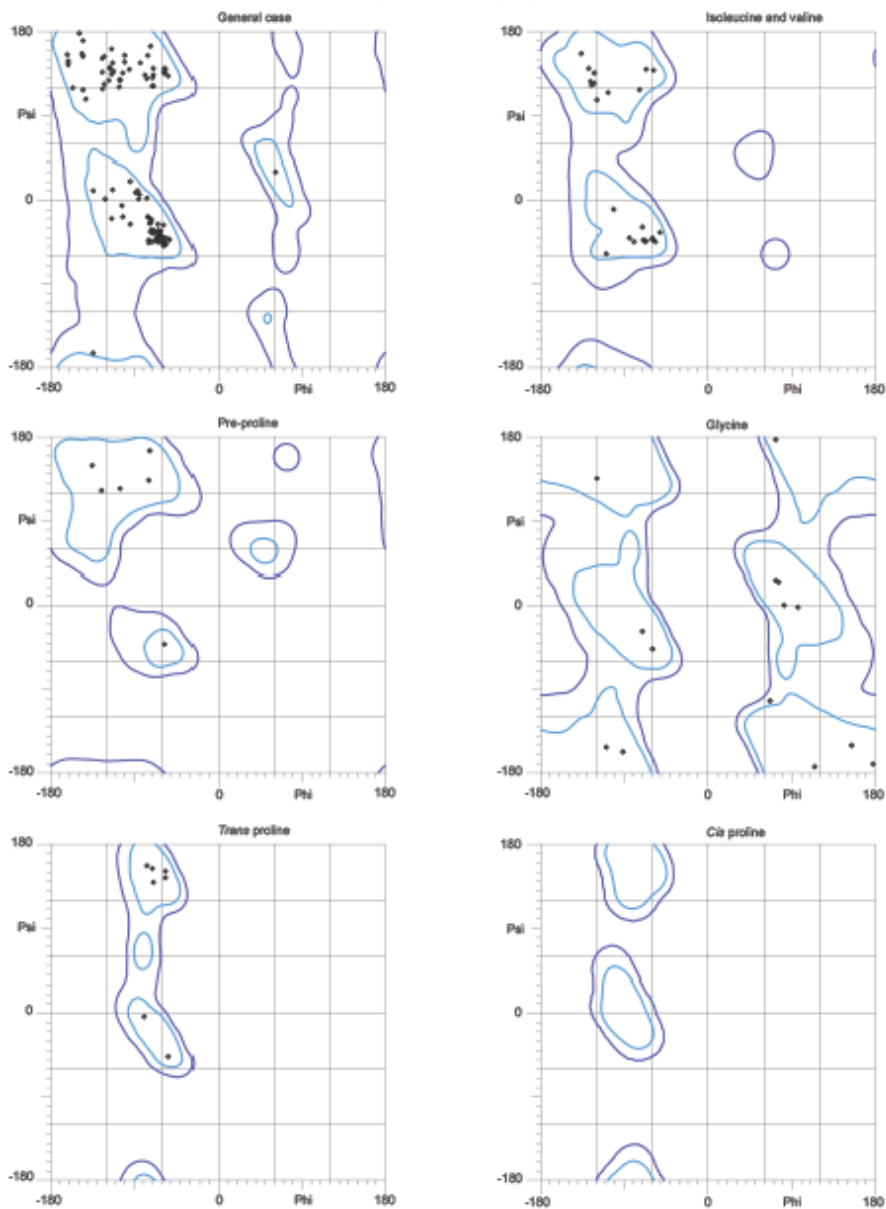


Figure 5.13: Electron density map of BPSL2418_{Met-SO} at 1.4 Å shows a Met-SO molecule binding the BPSL2418 close to CYS-75 and CYS-109 where a disulfide bond is formed between these two cysteines. Blue map calculated using coefficients $2mF_o-dF_c$ and contoured at 0.4σ . Green map calculated using coefficients mF_o-dF_c and contoured at 2.4σ .



98.8% (158/160) of all residues were in favored (98%) regions.
 100.0% (160/160) of all residues were in allowed (>99.8%) regions.

There were no outliers.

All-Atom Contacts	Clashscore, all atoms:	8.19	51 st percentile* (N=283, 1.20Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	0	0.00% Goal: <1%
	Ramachandran outliers	0	0.00% Goal: <0.05%
	Ramachandran favored	158	98.75% Goal: >98%
	MolProbity score [^]	1.44	78 th percentile* (N=1439, 1.20Å ± 0.25Å)
	Cβ deviations >0.25Å	1	0.68% Goal: 0

Figure 5.14: Result of MOLPROBITY and Romachandran plot of BPSL2418_{Met-SO} structure. (A) The Romachandran plot shows that all residues are within the favored region. (B) Overall the MOLPROBITY score is in the 78th percentile (100% being the best amongst structures of comparable resolution). MOLPROBITY (Chen *et al*, 2010).

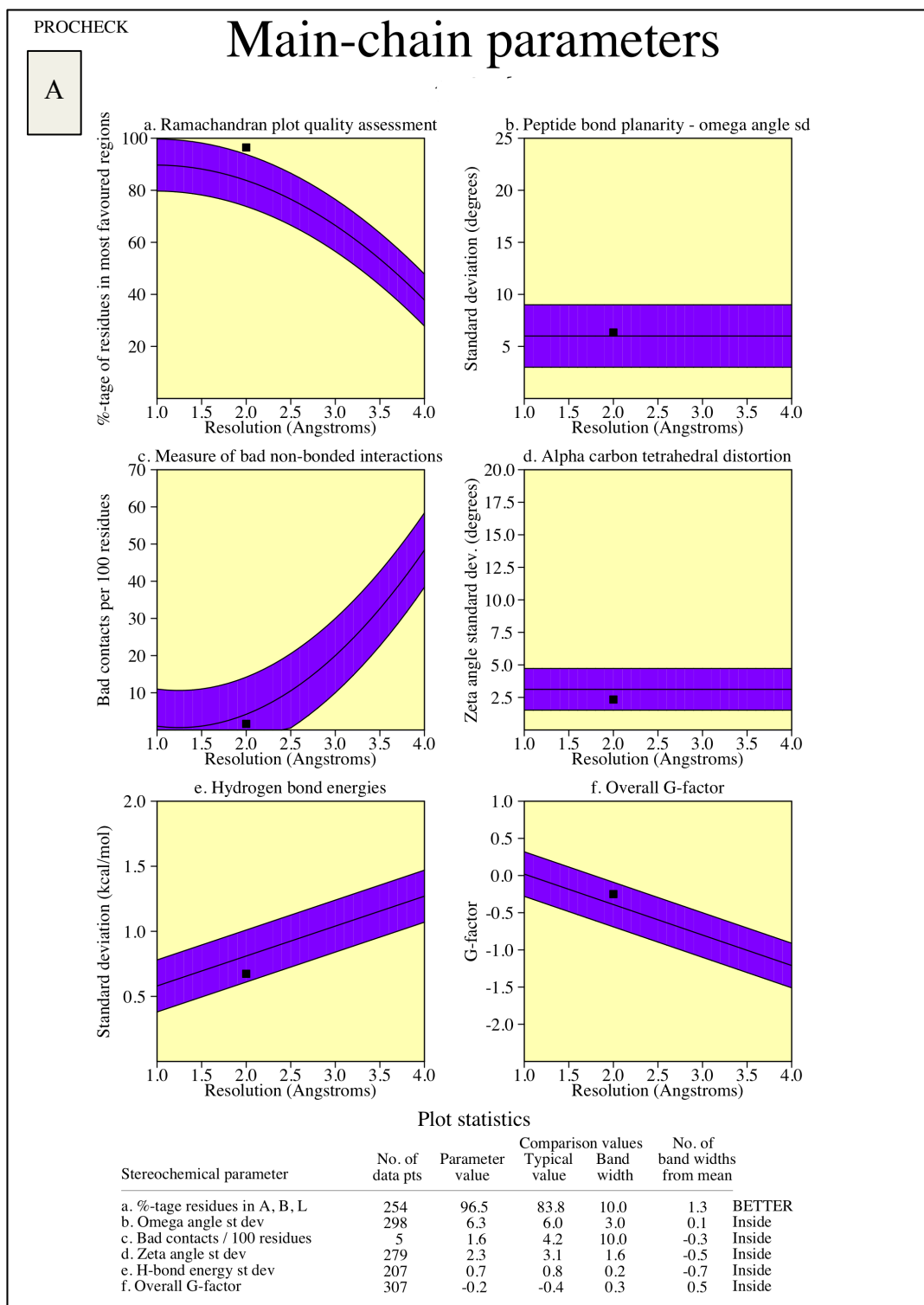
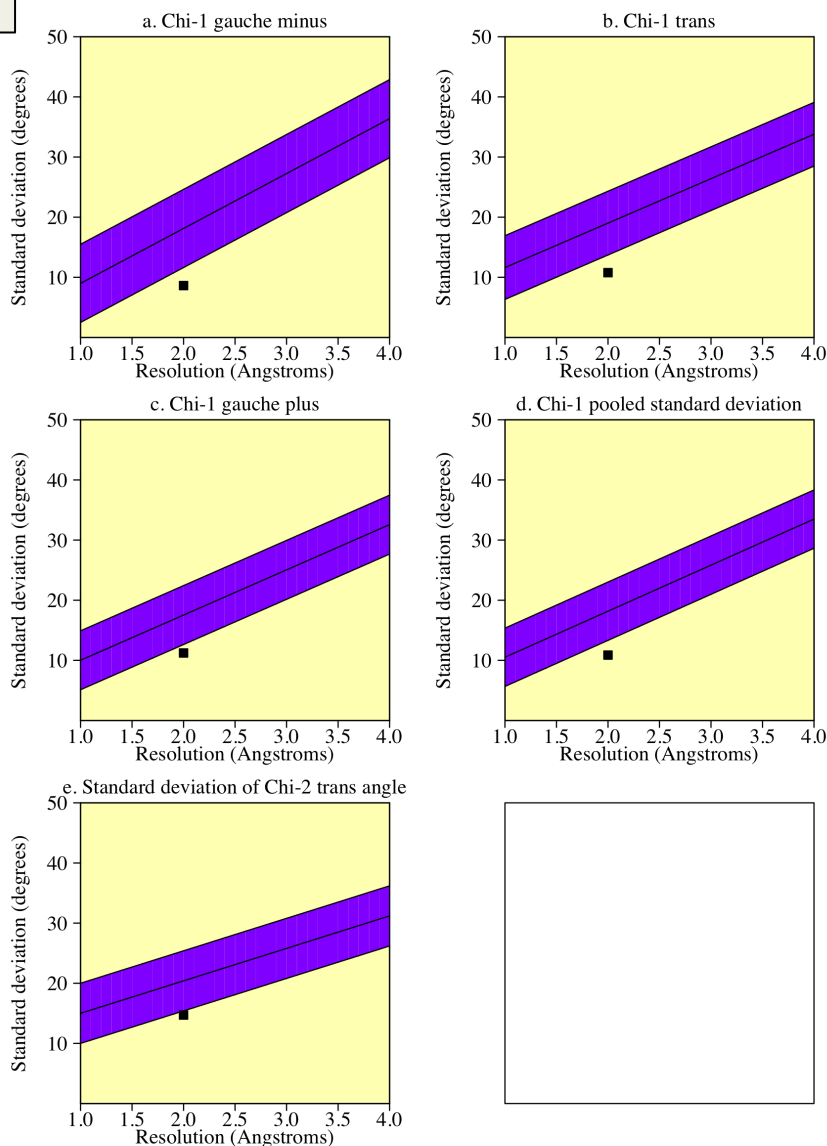


Figure 5.15: Main chain and side chain parameters of BPSL2418_{Met-SO}. (A) Showing all the main chain parameters values are better or within the average for a structure at this resolution.

Side-chain parameters

B

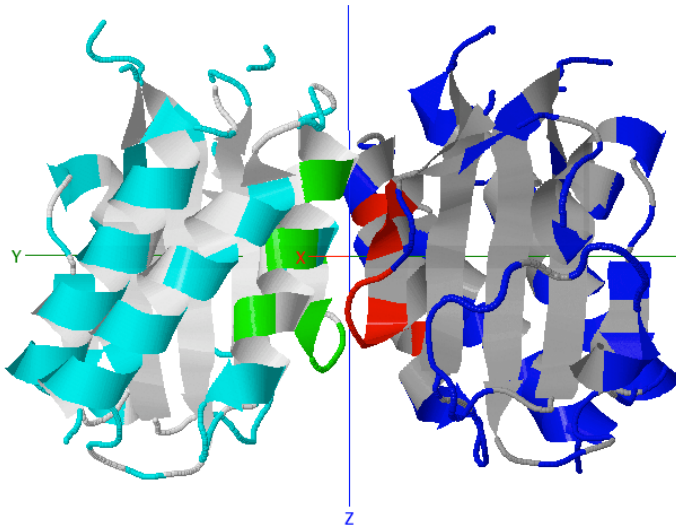


Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean	
			Typical value	Band width		
a. Chi-1 gauche minus st dev	25	8.6	18.1	6.5	-1.5	BETTER
b. Chi-1 trans st dev	88	10.8	19.0	5.3	-1.6	BETTER
c. Chi-1 gauche plus st dev	101	11.2	17.5	4.9	-1.3	BETTER
d. Chi-1 pooled st dev	214	10.9	18.2	4.8	-1.5	BETTER
e. Chi-2 trans st dev	55	14.7	20.4	5.0	-1.1	BETTER

(B) Showing all the side chain parameters values are better than the average for a structure at this resolution. Produced by the programme PROCHECK (Laskowski *et al*, 1993)

A



B

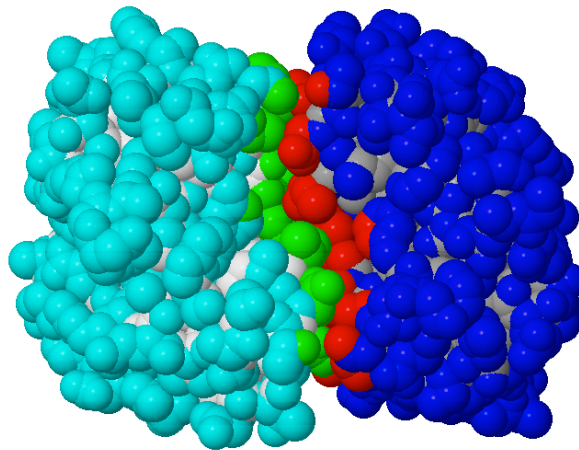


Figure 5.16: Dimeric BPSL2418_{Met-so}. (A) Cartoon model. (B) Space-filling model. The Figure was produced using PISA webserver http://pdbe.org/PISA_

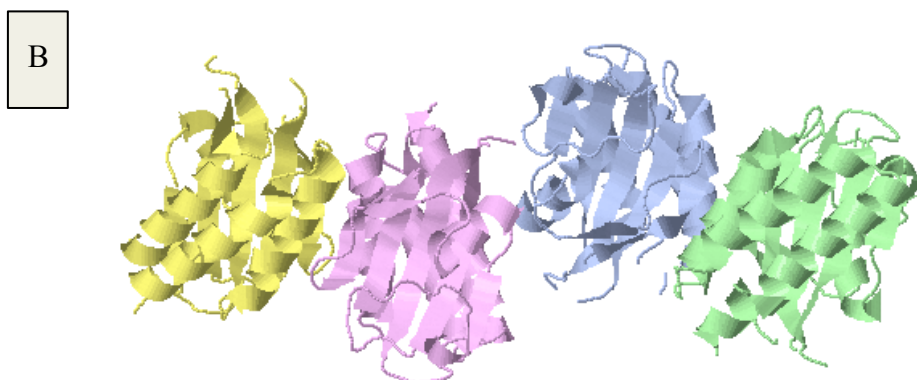
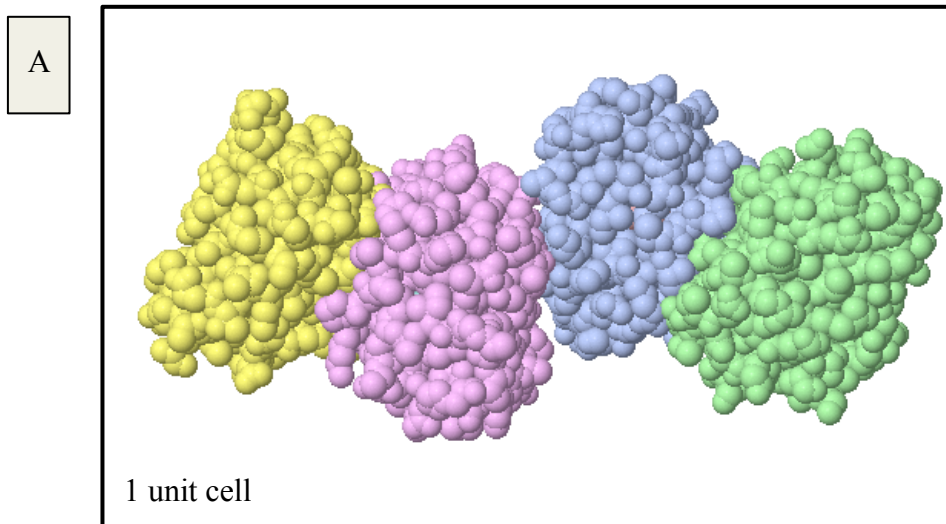


Figure 5.17: Two BPSL2418_{Met-SO} dimers in a unit cell (space group P2₁) of the crystal. (A) Spacefil model. (B) Cartoon model. The Figure was produced using PISA webserver <http://pdbe.org/PISA>.

5.3 Experimental structure determination of BPSL2418_{Reduced} by molecular replacement using BPSL2418_{MES}

The best crystals of reduced BPSL2418 were selected from the optimization trial for data collection (section 4.6.3).

5.3.1 Space group determination and indexing

The X-ray source in Sheffield was used for initial X-ray diffraction determination. The best diffraction of BPSL2418_{Reduced} was 3.1 Å and in order to achieve higher resolution diffraction the crystal was sent to the Diamond Light Source at the I02 beamline. To obtain a collection strategy, three initial test images 45° apart, 0.2° oscillation were collected using MOSFLM. BPSL2418_{Reduced} was indexed to the primitive monoclinic crystal system P2, with unit cell parameters: $a = 35.99 \text{ \AA}$, $b = 92.28 \text{ \AA}$, $c = 42.75 \text{ \AA}$, $\alpha = 90^\circ$, $\beta = 102.72^\circ$ and $\gamma = 90^\circ$; i.e. essentially isomorphous with BPSL2418_{Met-SO}.

5.3.2 Data collection and processing

The data from BPSL2418_{Reduced} crystal were collected using the I02 beamline of the Diamond Light Source. For the data set 900 images were collected with 0.2 oscillation per image using X-ray of wavelength 0.97949 Å and using a Pitatus 6M detector with a crystal-to-detector distance of 395.28 mm. Data were collected to 2.0 Å resolution (Figure 5.18). The images were processed using xia2 system and were indexed and integrated by XDS and scaled by XSCALE. All statistics of the processed data of BPSL2418_{Reduced} at 2.0 Å resolution are described in table 5.5. The processing gave overall $R_{\text{merge}} = 0.129$ (0.522 for the outer shell) and the $I/\sigma I$ was 6.3 overall (2.3 for the outer shell).

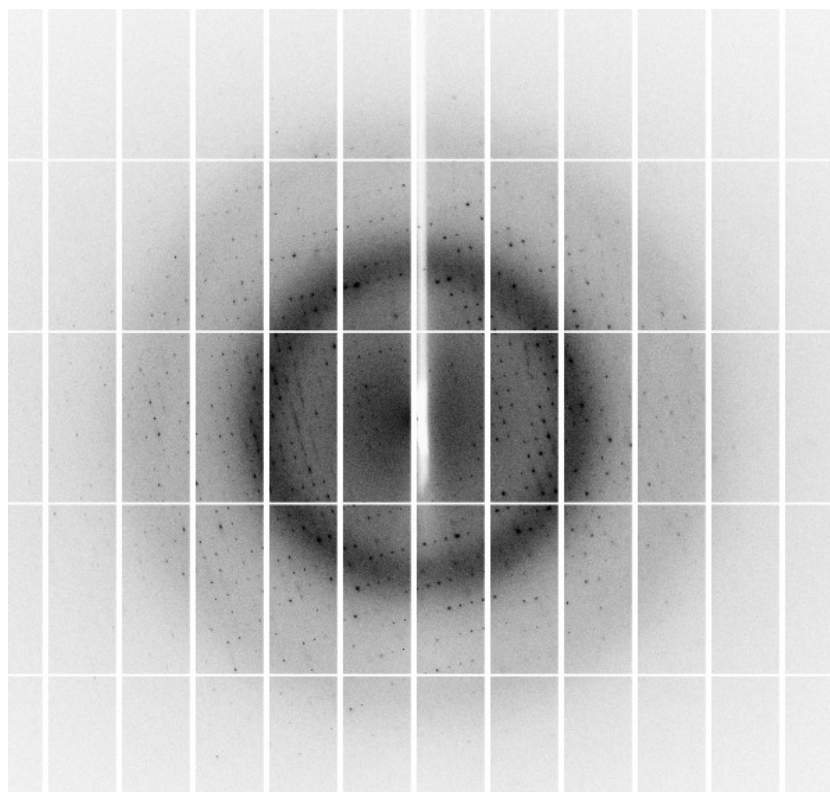


Figure 5.18: X-ray diffraction image of BPSL2418_{Reduced} crystal to 2.0 Å using the I02 beamline of the Diamond Light Source.

Data collection statistics	
Beamline at DLS,Oxford, UK	I02
Detector	Pitatus 6M
The crystal-to-detector distance	395.28 mm
Space group	P2 ₁
Crystal system	monoclinic
The unit cell parameters:	
a (Å)	35.99
b (Å)	92.28
c (Å)	42.75
α (°)	90.00
β (°)	102.72
γ (°)	90.00
Wavelength (Å)	0.97949
Resolution range (Å)	38-2.21 (2.27-2.21)
No. of unique reflections	13602 (998)
R_{merge}	0.13 (0.52)
R_{pim}	0.1 (0.41)
Mean (I)/σ(I)	6.3 (2.3)
Completeness (%)	99.4 (98.7)
Multiplicity	3.3 (3.1)
Mosaicity (°)	0.2
Number of molecules in the asymmetric unit	2

Table 5.5: Data collection statistics for the BPSL2418_{Reduced}. Number in parentheses shows the highest resolution shell.

5.3.3 Matthews Coefficient (V_m) calculation

The Matthews calculation for BPSL2418_{Reduced} using Mattprob webserver shows the possibility of one or two molecules in the asymmetric unit, with two as the most likely solution. The Matthews coefficient of the protein is 1.92 and estimated solvent content is 36.1% (Figure 5.19).

5.3.4 Structure determination of BPSL2418_{Reduced}

BPSL2418_{Reduced} structure was solved by molecular replacement using the model of BPSL2418_{MES} in PHASER. The 2.0 Å dataset of BPSL2418_{Reduced} were input and Phaser searched for two molecules in the asymmetric unit. A single solution was found using space group P2₁ giving rotation and translation function Z-scores of 10.1 and 32.7 respectively, which suggested the model was correct.

5.3.5 Model building and refinement

The obtained electron density map of BPSL2418_{Reduced} generated by molecular replacement was viewed in COOT. The initial model produced has an R-factor of 0.271 and R_{free} of 0.328. An improved model with a reduction in R-factor and R_{free} to 0.19 and 0.214, respectively was obtained by using repetitive cycles of model building and refinement (Figures 5.20 & 5.21). The refinement was carried on until there was no possible structure improvement. The final model consists of two molecules chain, A which consist of 156 residues, chain B with 155 residues and 116 water molecules.

5.3.6 3D structure of BPSL2418_{Reduced}

The crystal structure exhibits two molecules in the asymmetric unit of BPSL2418_{Reduced}, which agrees with the Matthews coefficient prediction. The electron density of these two molecules is well determined except 11 and 9 residues from the N termini and 3 and 6 residues from the C termini of chain A and chain B, respectively. The electron density map reveals the reduced state

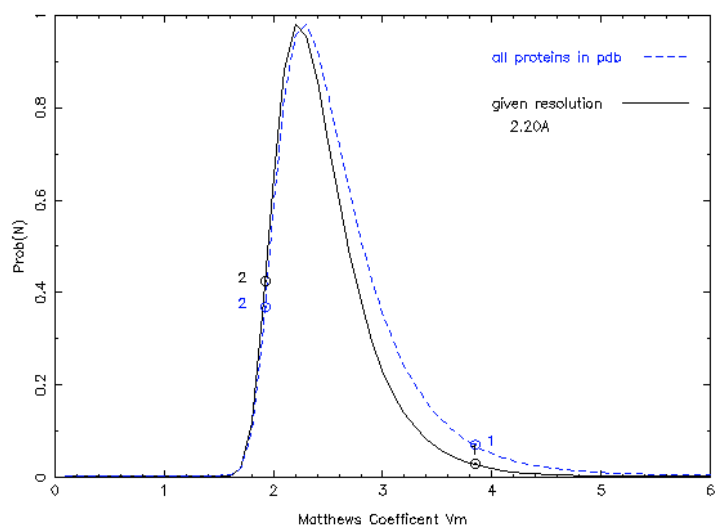
Found 2 possible solutions :

Vm of protein (Matthews coeff.) is 3.85 Å³/Dalton
 Estimated solvent content is 68.03 %

Vm of protein (Matthews coeff.) is 1.92 Å³/Dalton
 Estimated solvent content is 36.06 %

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm Å ³ /Da	Vs % solvent	Mw Da
1	0.0608	0.1582	3.85	68.03	18000.00
2	0.9392	0.8418	1.92	36.06	36000.00

B



C

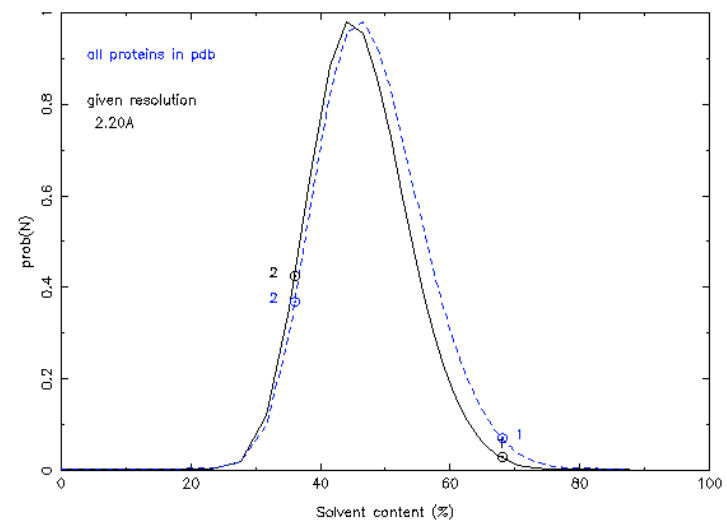


Figure 5.19: Matthews coefficient calculations and probabilities for BPSL2418_{Reduced} using Mattprob <http://www.ruppweb.org/Mattprob/>. This indicates the possibility of two protein molecules in the asymmetric unit (A) with a Vm of 1.92 Å³/Da (B) and a solvent content of 36.1 % (C).

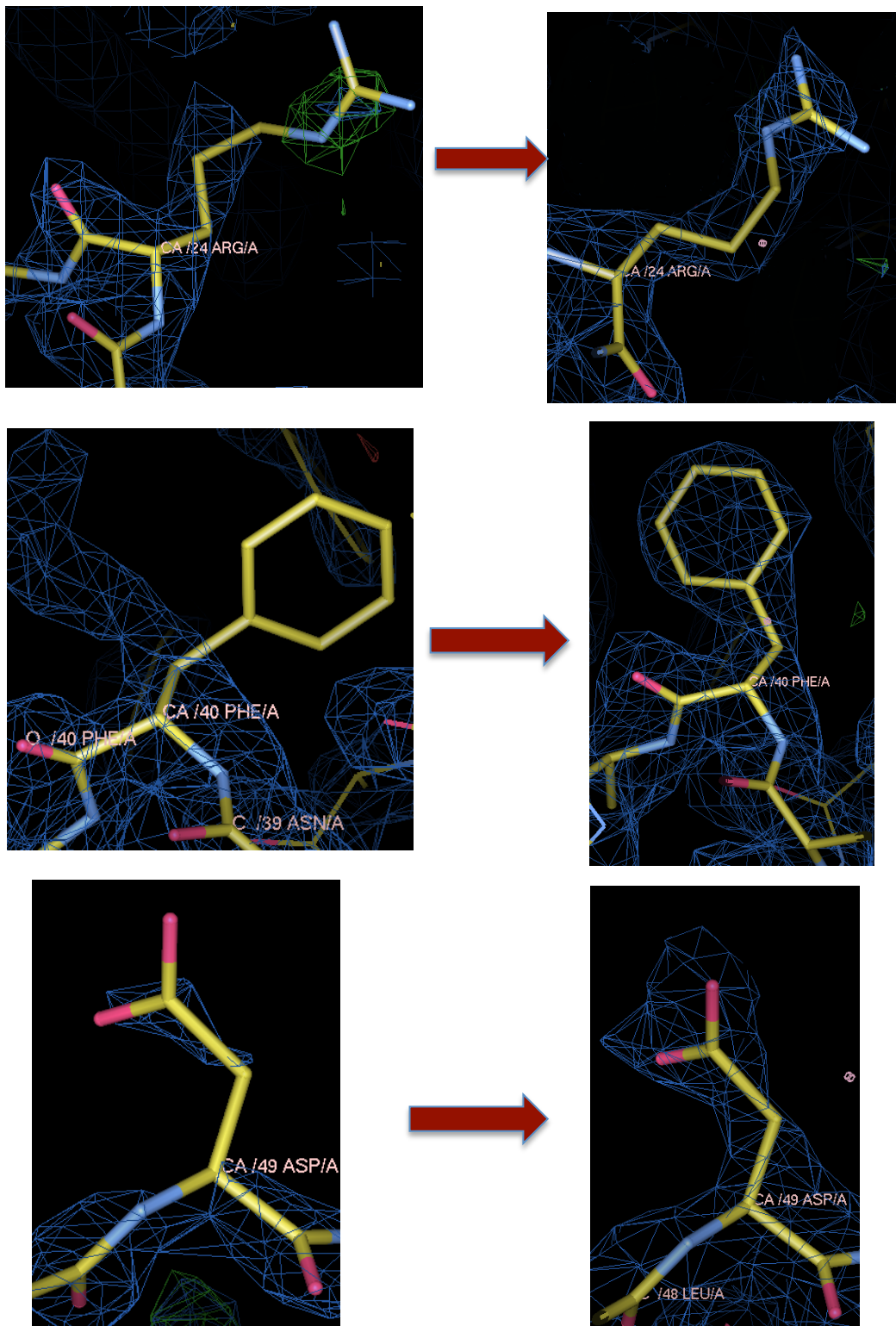


Figure 5.20: Sample regions of electron density for BPSL2418_{Reduced} chain A at 2.0 Å show the structure improvement over 18 refinement cycles. On the left three regions of electron density around residues 24, 40 and 49 after one cycle of refinement. On the right the same three regions after 18 refinement cycles. Blue map calculated using coefficients $2mFo-dFc$ and contoured at 0.69σ . Green map calculated using coefficients $mFo-dFc$ and contoured at 2.63σ .

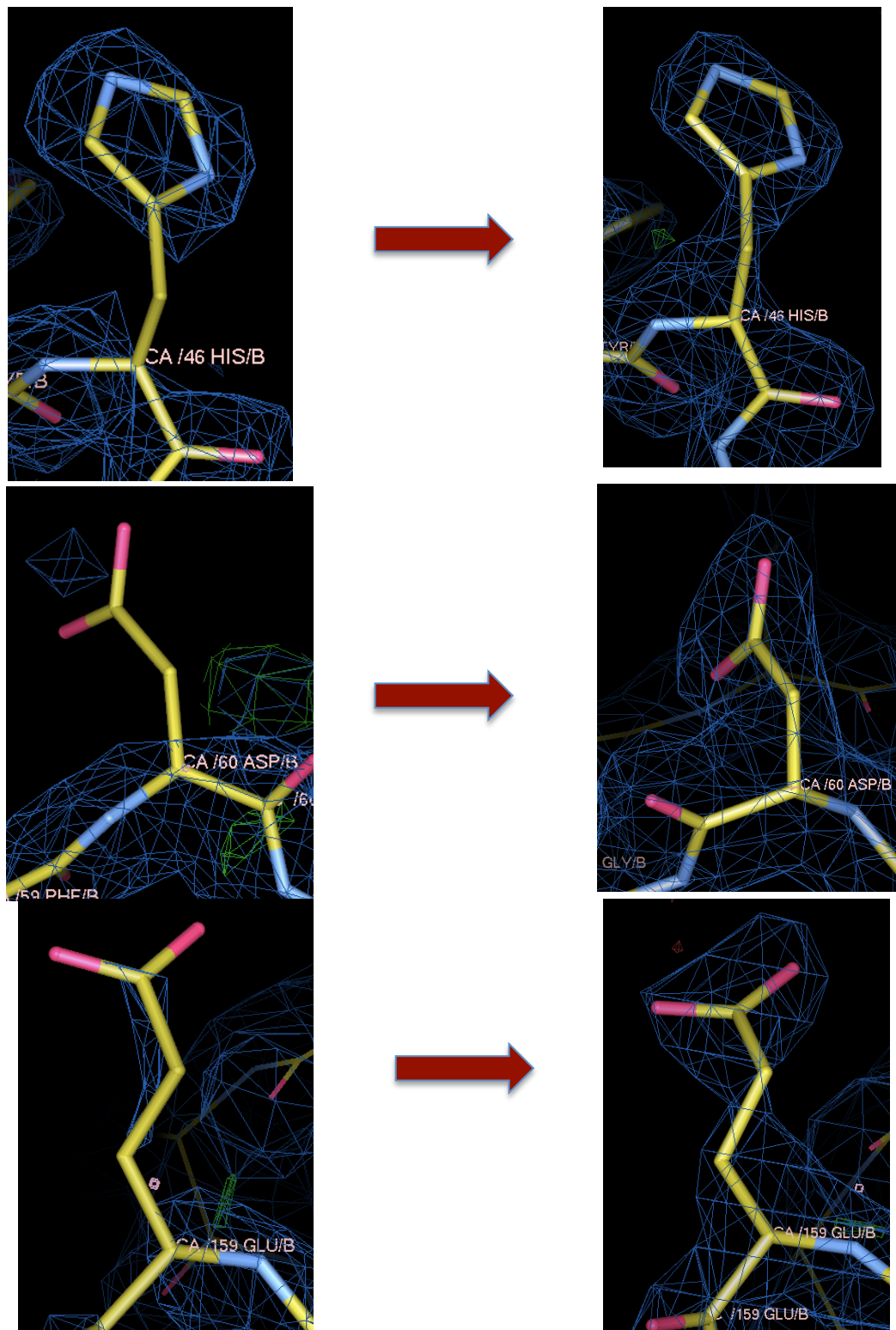


Figure 5.21: Sample regions of electron density for BPSL2418_{Reduced} chain B at 2.0 Å show the structure improvement over 18 refinement cycles. On the left side three regions of electron density around residues 46,60 and 159 after one cycle of refinement. On the right side the same three regions after 18 refinement cycles. Blue map calculated using coefficients $2mF_o-dF_c$ and contoured at 0.69σ . Green map calculated using coefficients mF_o-dF_c and contoured at 2.63σ .

of the active site of BPSL2418, where no disulfide bridge (S-S) is formed between CYS75 and CYS109 and the two cysteines are maintained in the sulphhydryl (-SH) form (Figure 5.22). BPSL2418_{Reduced} structure description and analysis will be discussed in chapter 6. The final model was validated using PROCHECK (Laskowski, 1993), and MOLPROBITY program (Chen *et al*, 2010). This shows that all residues fell within allowed regions of Ramachandran plot (Figure 5.23) and all main chain and side chain parameters were better or within the expected range for the resolution of data (Figure 5.24). These results reveal that the overall structure is of very good quality.

To determine the assemblies and interface between monomers in BPSL2418_{Reduced} the PISA webserver was applied. PISA was run using the reduced model of BPSL2418 which indicates a dimer interfaces between the two BPSL2418_{Reduced} monomers A and B (this is discussed in more details in chapter 6) (Figure 5.25). The BPSL2418_{Reduced} crystal belongs to space group P2₁ which has one 2-fold screw axis and this suggests that the unit cell contains two dimers of BPSL2418_{Reduced} (Figure 5.26).

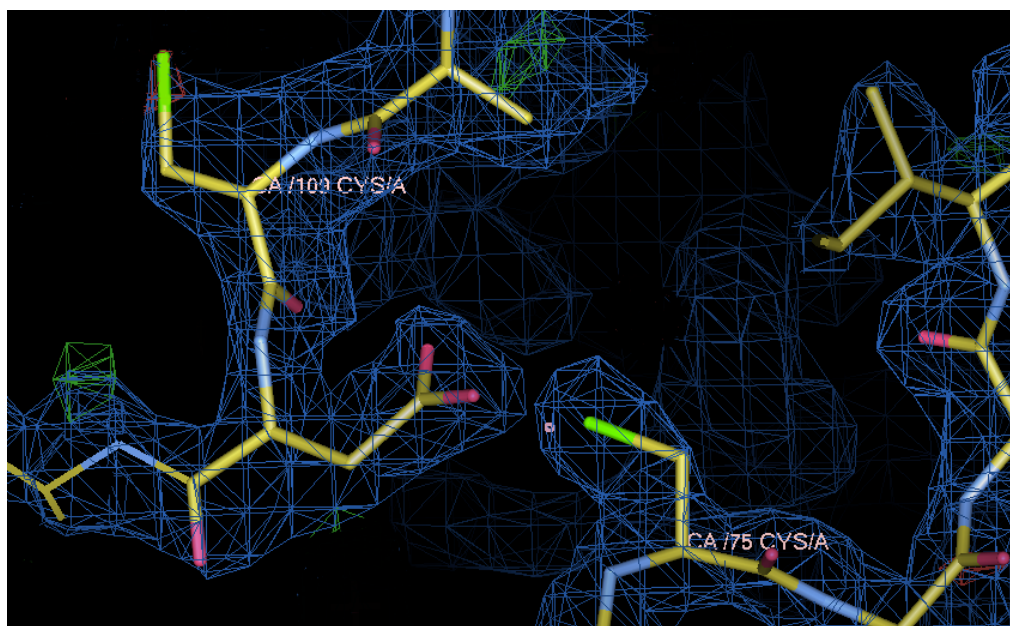
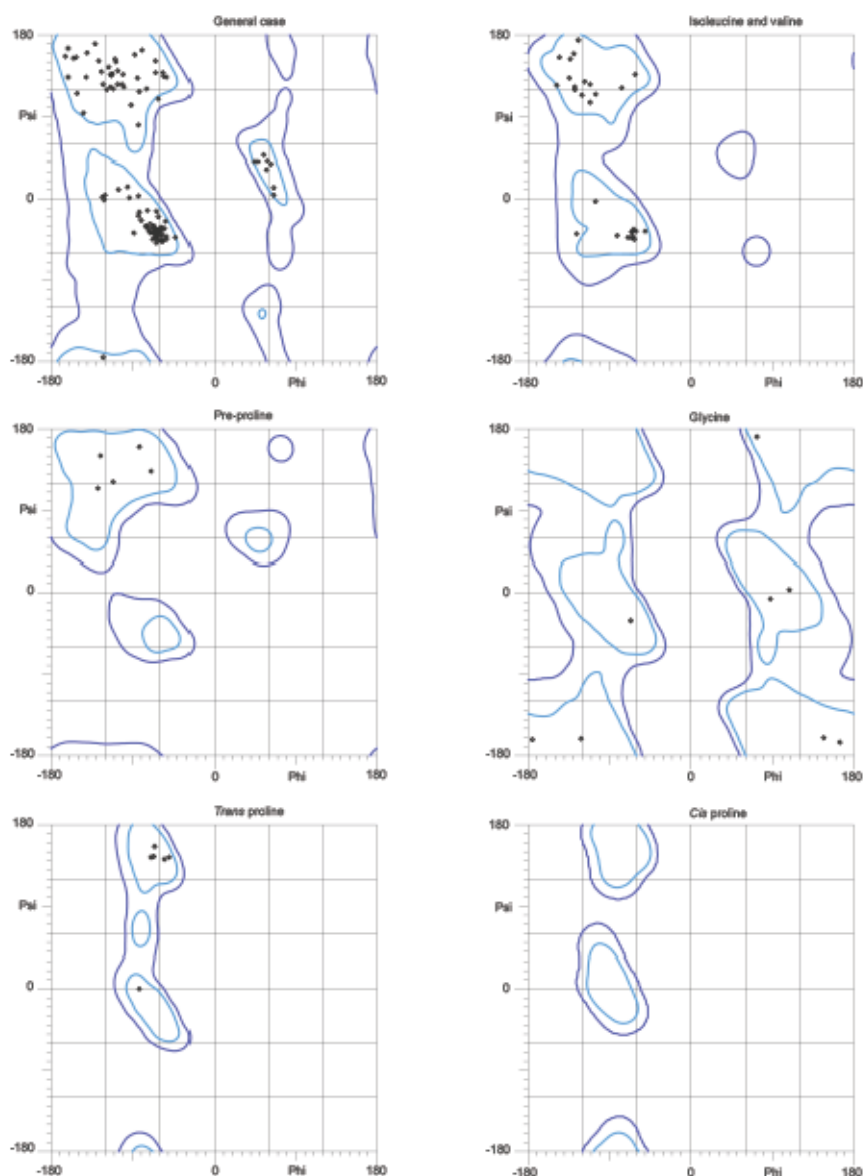


Figure 5.22: Electron density map of BPSL2418_{Reduced} at 2.0 Å shows no disulfide bond is formed between CYS75 and CYS109. Blue map calculated using coefficients $2mFo-dFc$ and contoured at 0.69σ . Green map calculated using coefficients $mFo-dFc$ and contoured at 2.63σ

Refinement statistics	
Resolution	2.0 Å
Protein molecules in asymmetric unit	2
Number of atoms	2403
Number of residues	311
Number of waters	116
Average B values (Å²):	
Whole chain A	27.2
Main chain A	26.7
Side chain A	27.8
Whole chain B	26.8
Main chain B	26.4
Side chain B	27.2
Water	44.3
R.m.s deviation:	
Bond lengths	0.0132 Å
Bond angles	1.7413°
Ramachandran plot:	
Most favored regions	98.0%
MolProbity score:	1.7/79 th percentile
R-factor	0.19
R_{free}	0.20

Table 5.6: Refinement statistics for BPSL2418_{Reduced} after 18 cycles of rebuilding in COOT and refinement in REFMAC5. The Ramachandran scores were produced using MolProbity.



98.0% (147/150) of all residues were in favored (98%) regions.
 100.0% (150/150) of all residues were in allowed (>99.8%) regions.

There were no outliers.

All-Atom Contacts	Clashscore, all atoms:	11.3	59 th percentile* (N=819, 1.70Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	2	1.52% Goal: <1%
	Ramachandran outliers	0	0.00% Goal: <0.05%
	Ramachandran favored	147	98.00% Goal: >98%
	MolProbity score ^a	1.71	79 th percentile* (N=9248, 1.70Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00% Goal: 0
	Bad backbone bonds:	0 / 1215	0.00% Goal: 0%
Bad backbone angles:	0 / 1648	0.00% Goal: <0.1%	

Figure 5.23: Result of MOLPROBITY and Ramachandran plot of BPSL2418_{Reduced} structure. (A) The Ramachandran plot shows that all residues are within the favored region. (B) Overall the MOLPROBITY score is in the 79th percentile (100% being the best amongst structures of comparable resolution). MOLPROBITY (Chen *et al*, 2010).

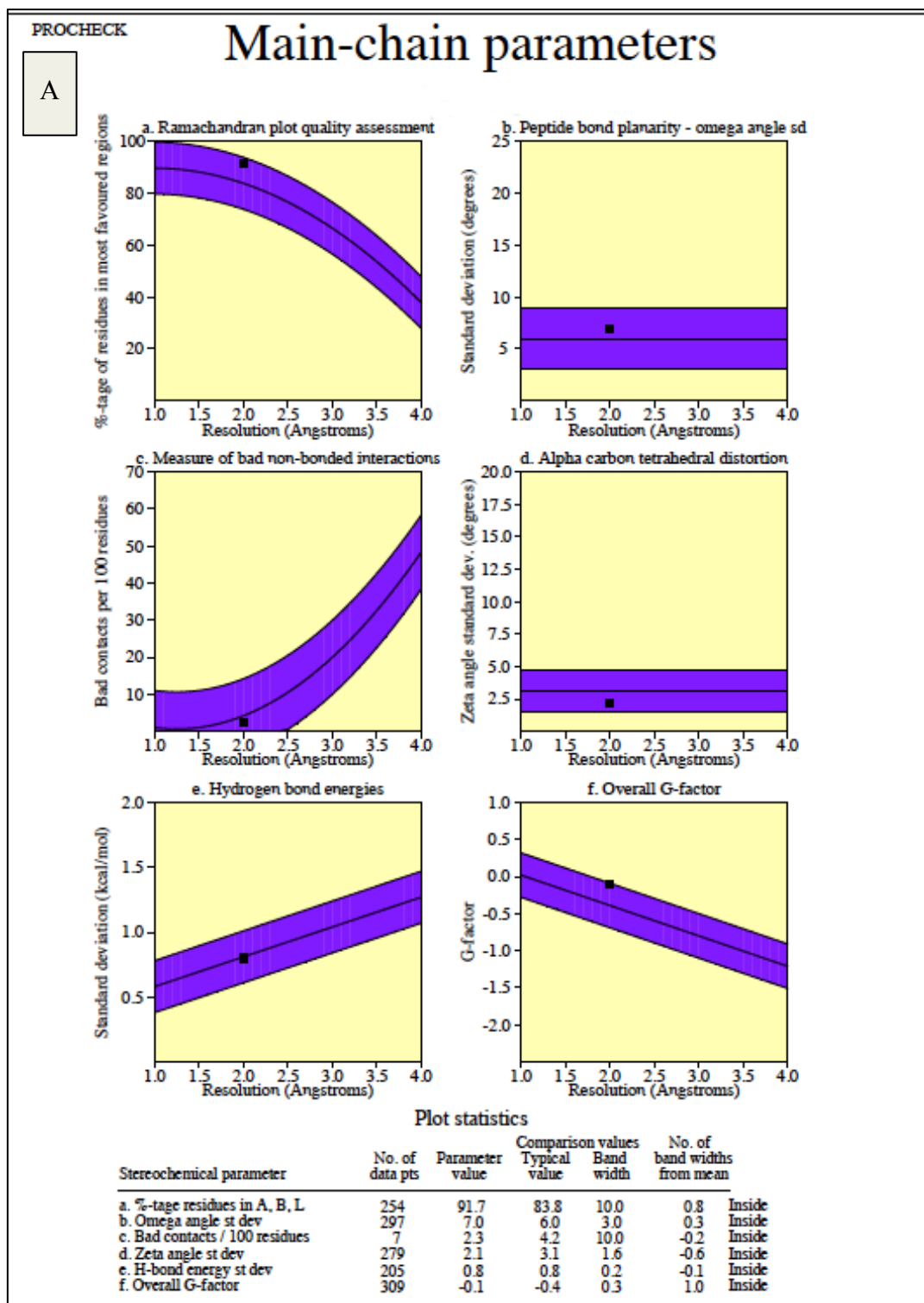
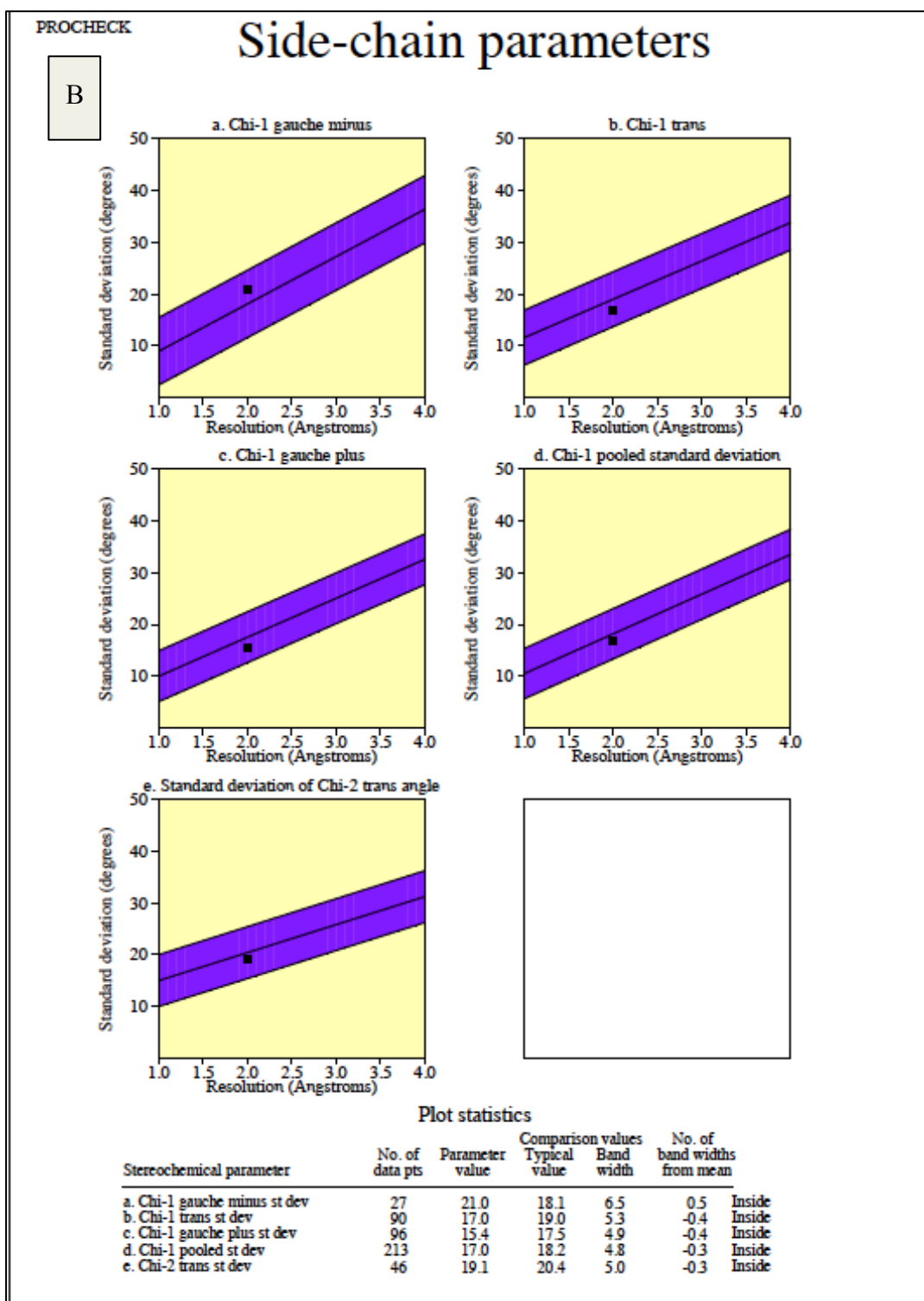


Figure 5.24: Main chain and side chain parameters of BPSL2418_{Reduced}. (A) Showing all the main chain parameters values are within the average for a structure at this resolution.



(B) Showing all the side chain parameters values are within the average for a structure at this resolution. Produced by the programme PROCHECK (Laskowski *et al*, 1993)

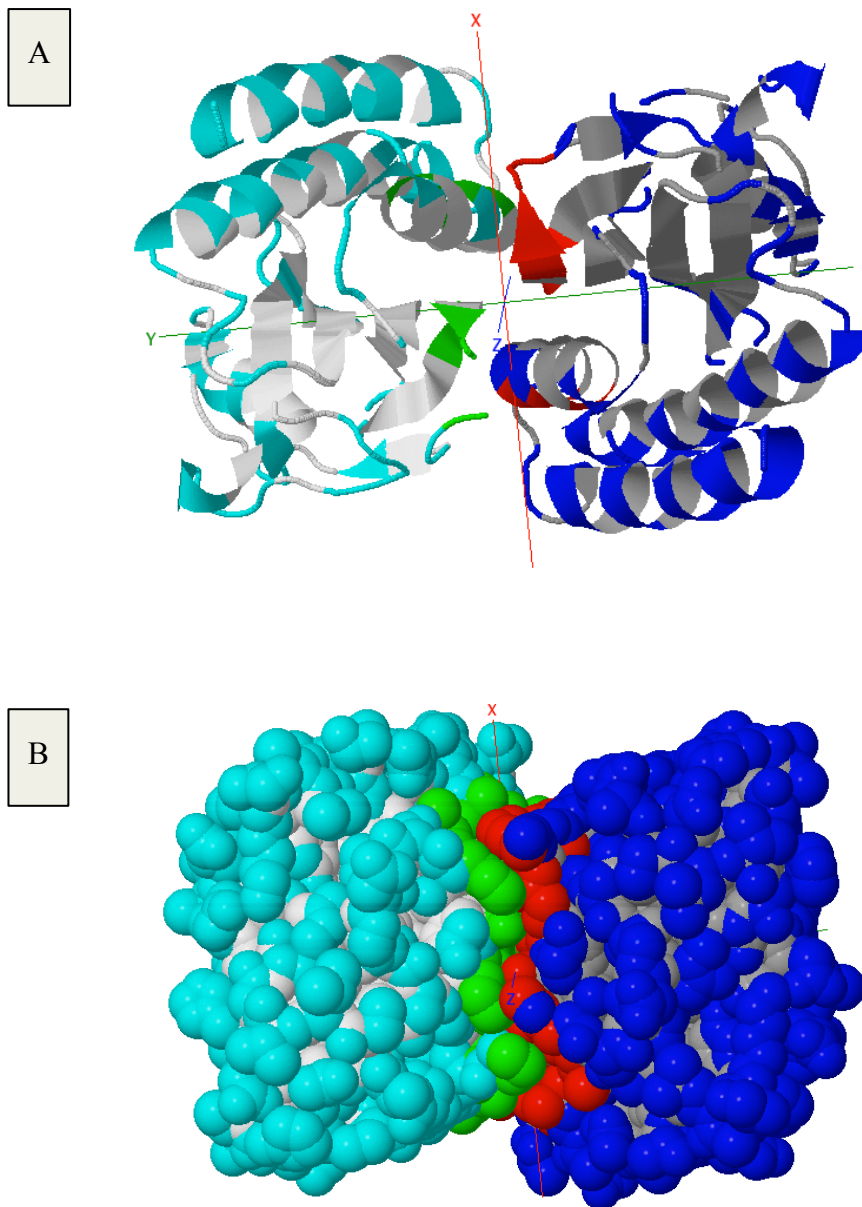


Figure 5.25: Dimeric BPSL2418_{Reduced}. (A) Cartoon model. (B) Space-filling model. The Figure was produced using PISA webserver http://pdbe.org/PISA_

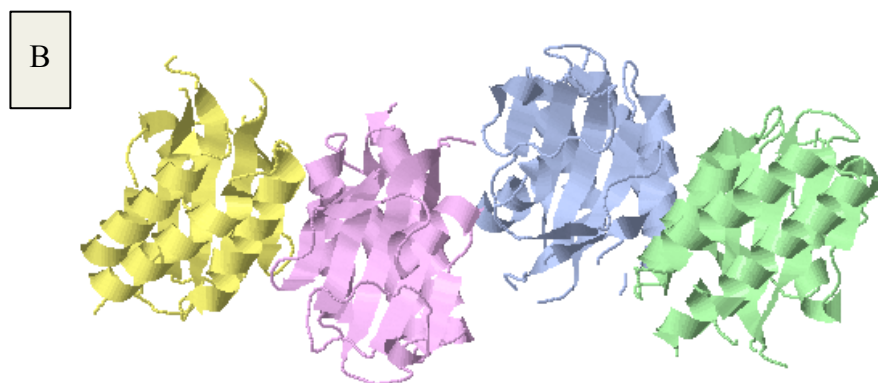
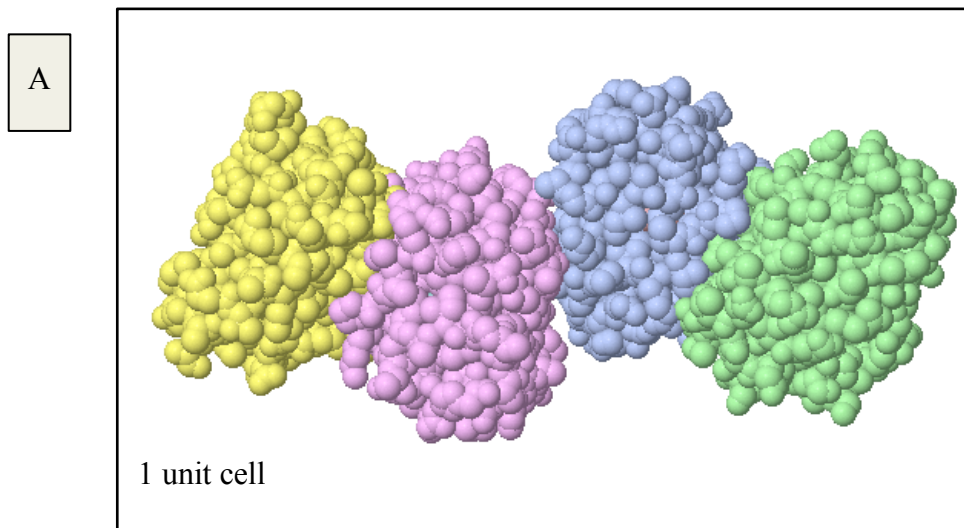


Figure 5.26: Two BPSL2418_{Reduced} dimers in a unit cell (space group P2₁) of the crystal. (A) Spacefill model. (B) Cartoon model. The Figure was produced using PISA webserver <http://pdbe.org/PISA>.

Chapter 6: BPSL2418 structure analysis

This chapter represents the analyses of the structures of MES-bound, Met-SO bound and reduced forms of BPSL2418, which were produced in this project. Also the chapter describes the putative active site and BPSL2418 is possible role in *Burkholderia pseudomallei*.

6.1 BPSL2418 structure description

This section describes in detail the structure analysis of BPSL2418_{MES}, BPSL2418_{Met-SO} and BPSL2418_{Reduced}.

6.1.1 Features of BPSL2418 structure

6.1.1.1 BPSL2418_{MES}

The structure of BPSL2418_{MES} was determined from crystals grown in orthorhombic spacegroup the $P 2_1 2_1 2_1$ and it comprises one subunit in the asymmetric unit. A Dali search suggests that the BPSL2418_{MES} subunit structure shares the overall topology of the GAF domain and it consists of six antiparallel β -strands ($\beta 1$ - $\beta 6$) and four α -helices ($\alpha 1$ - $\alpha 4$) with connecting loops (Figure 6.1). The antiparallel β -strands ($\beta 1$, $\beta 2$, $\beta 5$ and $\beta 6$) are located in the center of the BPSL2418_{MES} structure flanked on one side by three α -helices ($\alpha 1$, $\alpha 2$ and $\alpha 4$). On the other side a construct of loop1($\beta 2\beta 3$ loop), $\beta 3$, loop2($\beta 3\alpha 3$ loop), $\alpha 3$, loop3($\alpha 3\beta 4$ loop), $\beta 4$ and loop4($\beta 4\beta 5$ loop) unit, where loop 4 forms a capping flap, which covers the active site (Figure 6.2 a).

The Active Site: The active site consists of residues from the four β -strands ($\beta 1$, $\beta 2$, $\beta 5$, $\beta 6$) and the loop1- $\beta 3$ - loop2- $\alpha 3$ - loop3- $\beta 4$ - loop4 unit that surrounds the active site forming a small cavity (Figure 6.2b). It contains TRP53 and TYR57 from $\beta 1$, ALA74 and CYS75 from loop1, ILE 78 from $\beta 3$, GLY83 from loop2, VAL84 and CYS85 from $\alpha 3$ and ILE107, ALA108, CYS109, SER 111, SER113 from loop4 (Figure 6.3b).

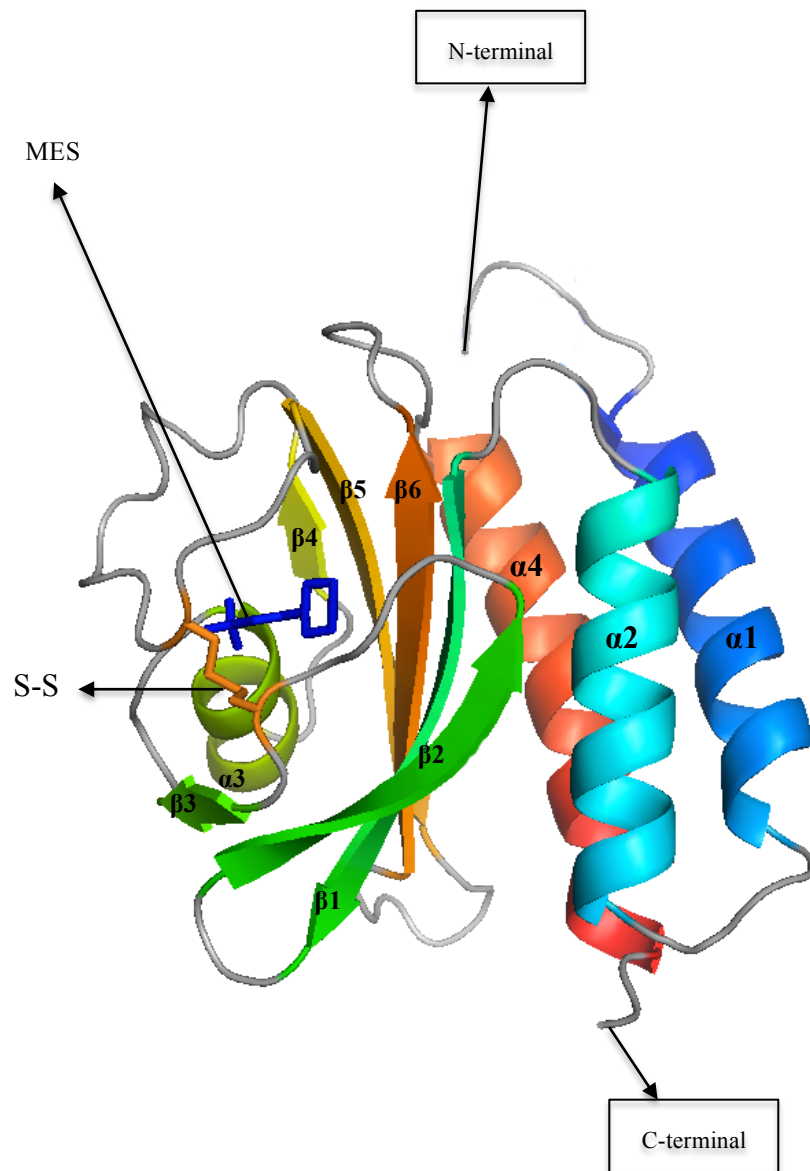


Figure 6.1: Cartoon representation of the overall fold structure of BPSL2418_{MES}. The structure consists of six β -strands and four α -helices with connecting loops. The disulfide S-S bond between CYS75 and CYS109 is shown in orange color and MES molecule in blue color. The Figure was created by PyMol (DeLano & Lam, 2005).

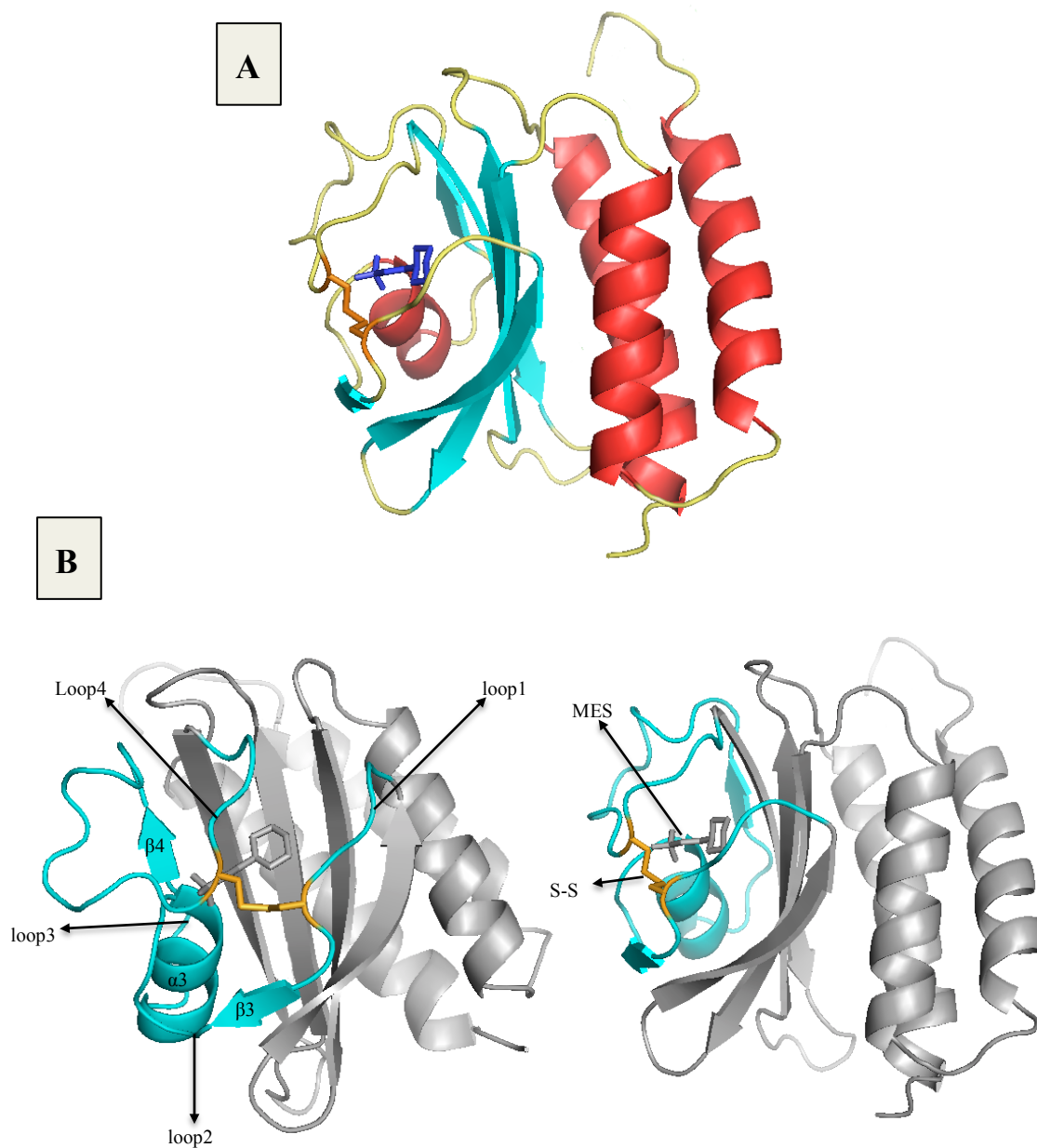


Figure 6.2: Cartoon representation of the BPSL2418_{MES} showing the active site. (A) The β -strands ($\beta 1$, $\beta 2$, $\beta 5$ and $\beta 6$) of BPSL2418_{MES} are located in the center of the structure flanked on one side by three α -helices ($\alpha 1$, $\alpha 2$ and $\alpha 4$), and on the other side by loop1, $\beta 3$, loop2, $\alpha 3$, loop3, $\beta 4$ and loop4 construct. (B) loop4 forms a capping flap, which covers the active site. The Figures were created by PyMol (DeLano & Lam, 2005).

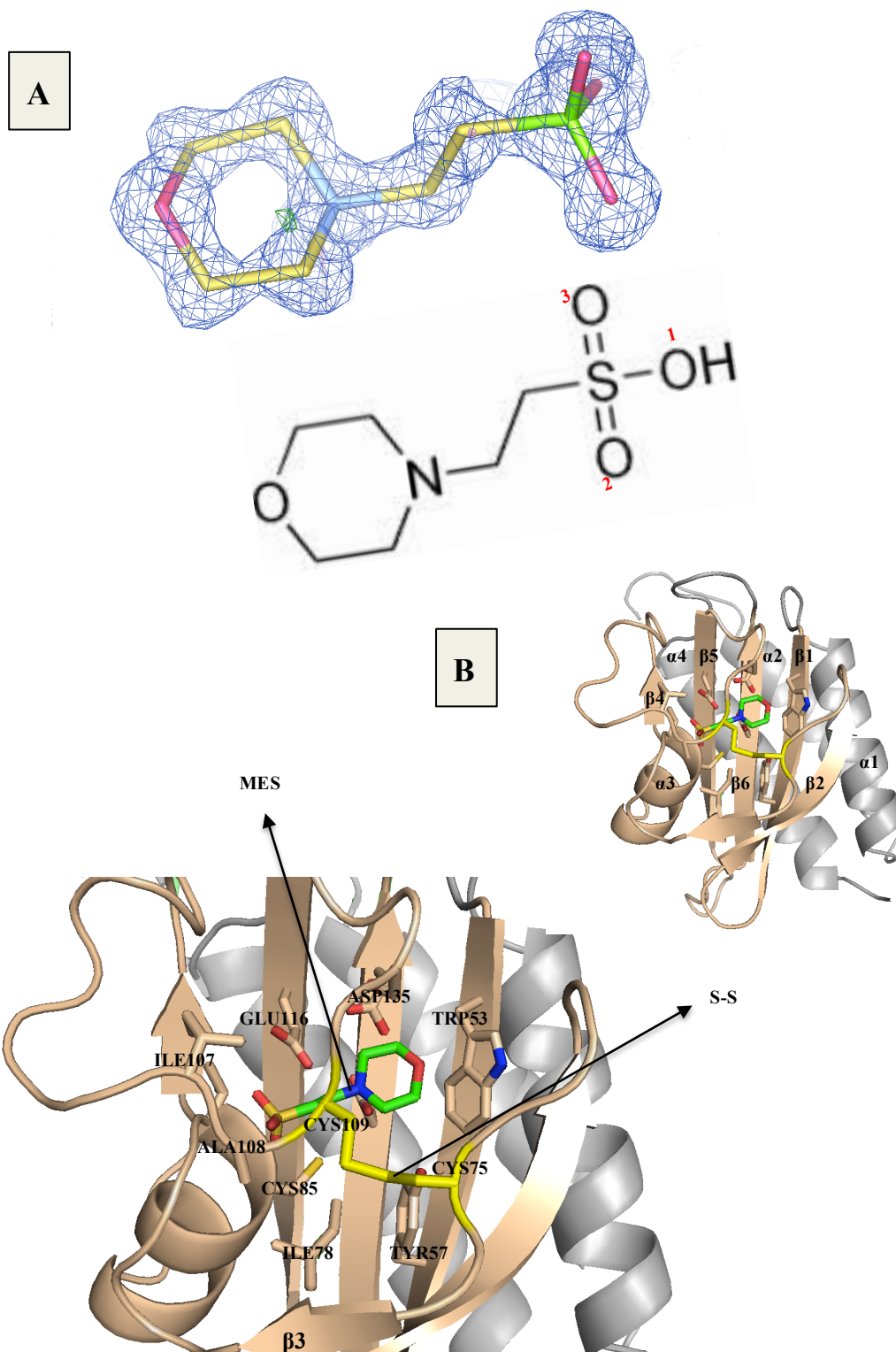


Figure 6.3: (A) Electron density map of 2-(N-morpholino)ethanesulfonic acid [MES] structure at 1.18 Å. (oxygen atom in red, nitrogen in blue and sulfur in green) (B) The binding site of BPSL2418_{MES} molecule with the MES bound. Residues involved in BPSL2418-MES binding are shown as sticks. The Figures were created by PyMol (DeLano & Lam, 2005). Blue map calculated using coefficients $2mF_o - dF_c$ and contoured at 0.5σ .

BPSL2418_{MES} structure is an oxidized form where a disulfide bond is formed between CYS75-CYS109, binding loop1 and loop4 together and tightening the active site cavity (Figure 6.3b). The MES molecule (Figure 6.3a) lies inside the cavity with a MES sulfonic acid group being close to CYS85 in $\alpha 3$ (Figure 6.3b). To analyze the interaction between MES and the active site of BPSL2418 protein, LigPlot (Wallace *et al*, 1995) was used. Hydrogen bonds are formed between O1 of the sulfonic acid group and the peptide nitrogen atoms of ALA108 and CYS109 residues, which are both located in loop4. Further hydrogen bonds are formed between O2 and O3 of the sulfonic acid group and the NH of VAL84 and CYS85 respectively, where VAL84 and CYS85 are both situated at $\alpha 3$. TYR57 and TRP53 were thought to provide the hydrophobic contact to the morpholine ring of MES, where both residues are situated in $\beta 1$. CYS85 and VAL84 in $\alpha 3$ and ALA108 in loop4 were proposed to form hydrophobic interactions with the sulfonic acid group of MES. In addition to ILE78, GLY83, ILE107 and SER111 residues, which predicted to form a hydrophobic pocket in the active site seems important in ligand binding (Figure 6.4).

Similar 3D structures: A search of the Dali database (Holm & Rosenstrom, 2010) with the BPSL2418_{MES} structure indicates several homologs with high Z-scores and high sequence identity. It shows that BPSL2418_{MES} belongs to a GAF domain family that acts as free methionine sulfoxide reductase, as also has been predicted before from protein sequence (see section 1.2.4). Dali suggests several homologs to BPSL2418_{MES} (table 6.1), but the top of the list is the *E. coli* fRMsR structure (PDB code 1VHM) (Badger *et al*, 2005) (Figure 6.5). Dali superimposed the BPSL2418_{MES} and the *E. coli* fRMsR over 151 α -carbon atom pairs with a Z-score of 26.8, sequence identity 49% and a root mean-square deviation (RMSD) of 1.1 Å. The *E. coli* fRMsR structure consists of two chains A and B, both have high identity to BPSL2418_{MES}. Similar to BPSL2418_{MES}, the *E. coli* fRMsR subunits have a GAF domain topology composing of four six antiparallel β -strands and four α -helices with connecting loops. BPSL2418_{MES} and the *E. coli* fRMsR each have three cysteines and a disulfide bond in the active site pocket, which are conserved in all of the fRMsR family (see section 1.2.2.3 and Figure 1.12). Also the *E. coli* fRMsR structure has a MES ligand bound in the active site with a similar set of residues

Number	PDB-Chain	Z-score	R.M.S.D	Aligned residues	Length of PDB model	Sequence identity (%)	Molecule description
1	1vhm-A	26.8	1.1	151	159	49	Structure of fRMsr of <i>E. coli</i> (Badger <i>et al</i> , 2005) Method: X-ray diffraction Resolution: 2.1 Å Ligand: MES
2	1vhm-B	26.6	1.3	152	160	49	Structure of fRMsr of <i>E. coli</i> (Badger <i>et al</i> , 2005) Method: X-ray diffraction Resolution: 2.1 Å Ligand: MES
3	3 ksf-G	25.7	2.1	151	152	44	Structure of fRMsr of <i>Staphylococcus aureus</i> (Bong <i>et al</i> , 2010) Method: X-ray diffraction Resolution: 1.9 Å Ligand: PEG
4	3 rfb-A	25.6	1.2	151	161	40	Structure of fRMsr of <i>Streptococcus pneumonia</i> (To be published) Method: X-ray diffraction Resolution: 2.3 Å Ligand: Met-R-SO
5	3 ksh-A	25.6	1.8	151	152	48	Structure of fRMsr of <i>Staphylococcus aureus</i> (Bong <i>et al</i> , 2010) Method: X-ray diffraction Resolution: 1.5 Å Ligand: SO ₄

Table 6.1: Dali server results for the model of BPSL2418_{MES}. The top five hits are listed alongside their related Zscores, RMSD scores, alignment statistics and a brief description of each protein.

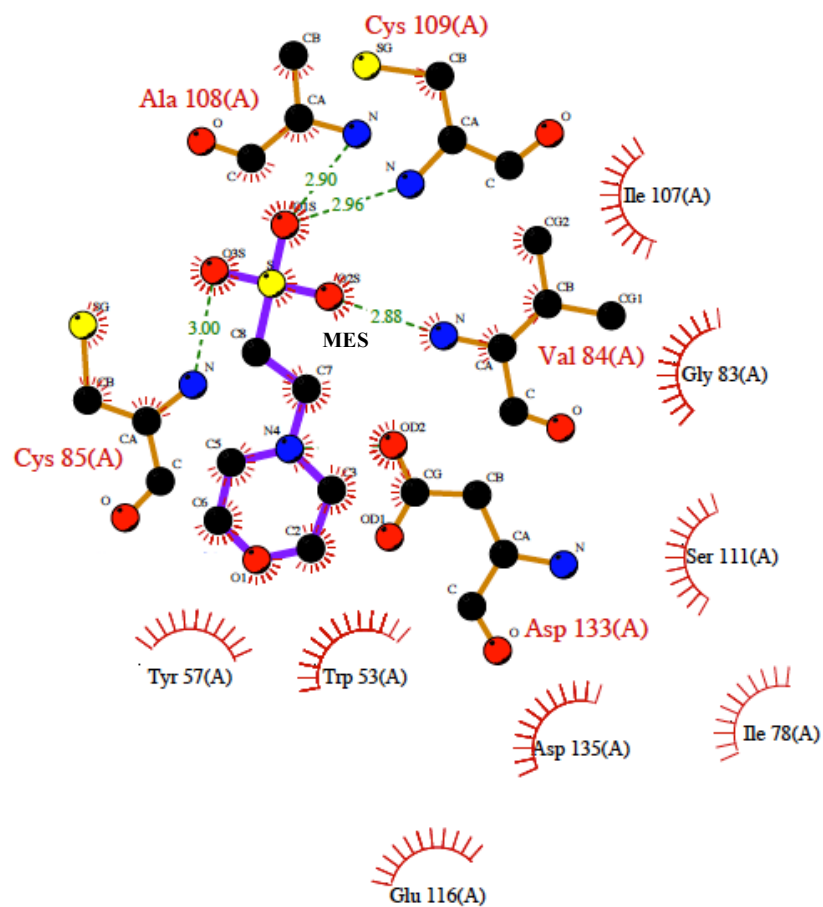


Figure 6.4: The binding site interaction of BPSL2418_{MES} with the MES buffer. The Figure shows the residues, are involved in hydrogen bonds and hydrophobic interactions. Hydrogen bonds are shown as green dashed lines between atoms. Hydrophobic interactions are indicated by a red lashes-like symbol pointing towards the ligand atom they contact. The MES atoms that are involved in hydrophobic interaction are represented by a red spokes pointing back. (oxygen atom in red, nitrogen in blue, carbon in black, sulfur in yellow). This Figure was produced by LigPlot (Wallace *et al*, 1995).

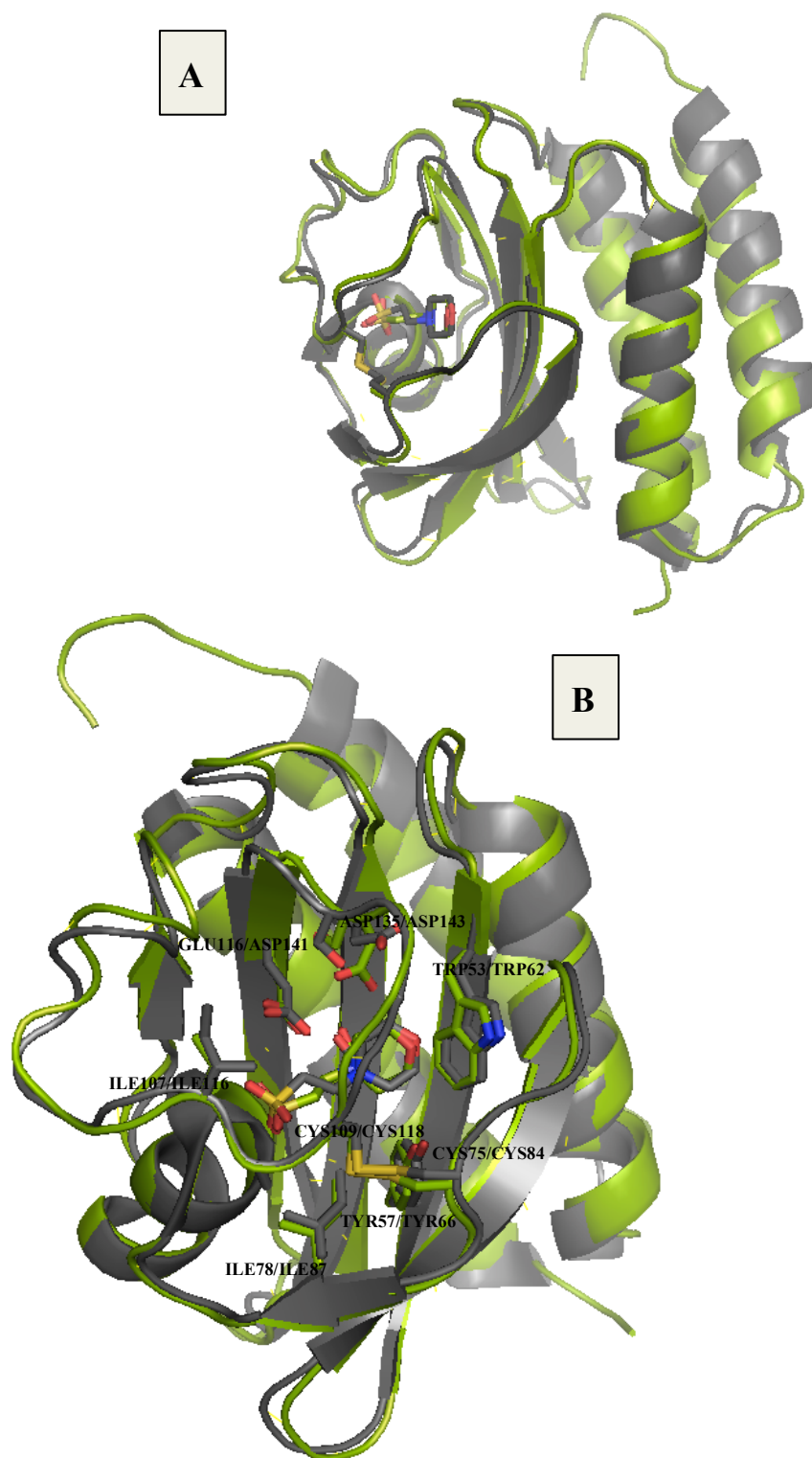


Figure 6.5: Superimposition of BPSL2418_{MES} (green) and *E. coli* fRMsr (gray). (A) Overall fold of BPSL2418_{MES} and *E. coli* fRMsr. (B) The active sites of the two structures showing the binding residues and MES ligand as sticks. It is clear that the overall structures and the active sites of both structures are almost identical. The Figure produced by PyMol (DeLano & Lam, 2005).

providing hydrogen bonds and hydrophobic surface (Figure 6.6). Similar to BPSL2418_{MES}, the sulfonic acid group of the MES ligand in *E. coli* fRMs_r active site is situated close to CYS94, the corresponding residue to CYS85 in BPSL2418_{MES}. The proximity of the MES sulfonic acid group to the third cysteine, which does not participate in the disulfide bond, is constant in the arrangement of the active site residues in the MsrA family and in the *E. coli* fRMs_r the only known structure from the fRMs_r family that has a MES molecule in the active site (Lin *et al.*, 2007).

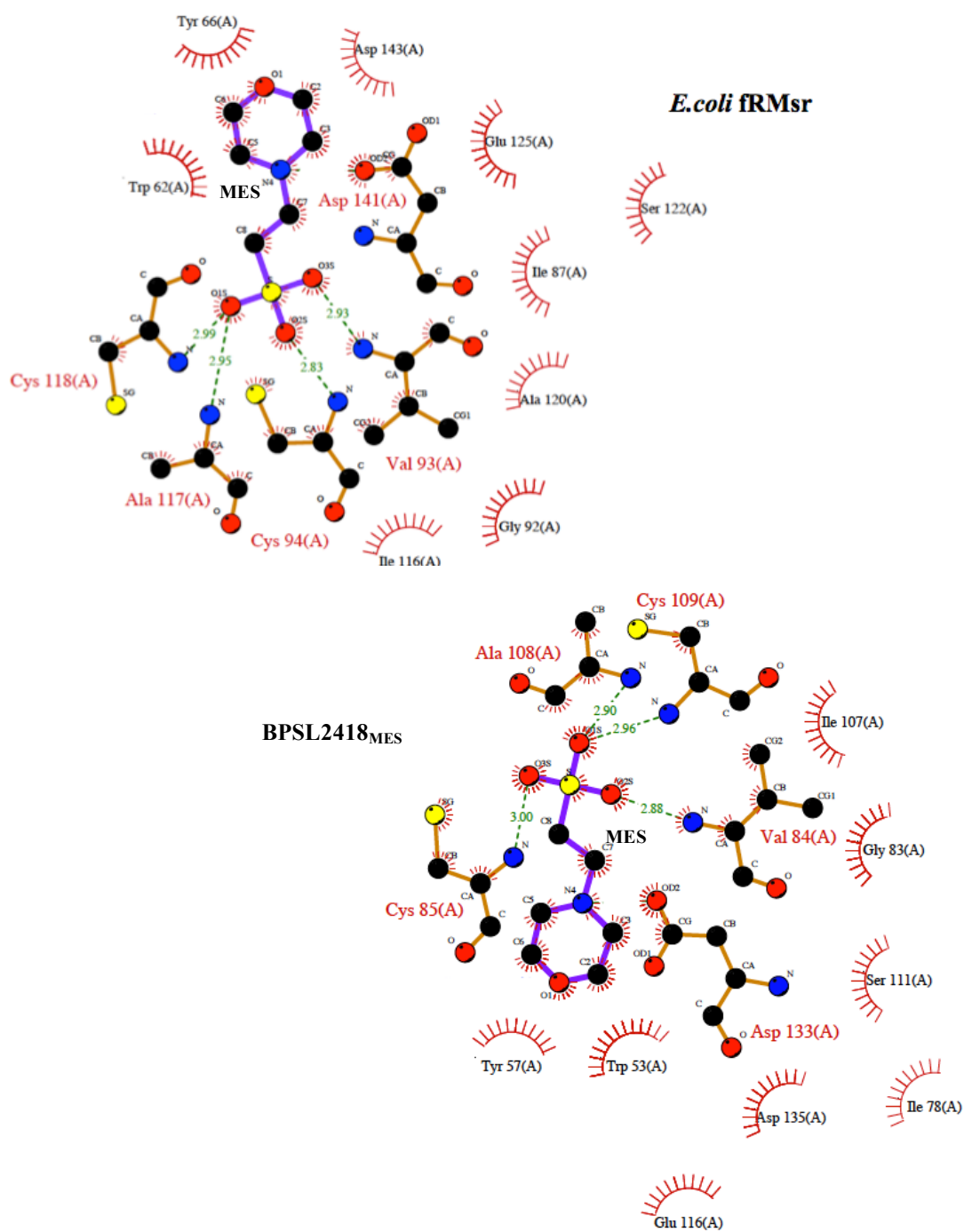


Figure 6.6: LigPlot diagrams of the binding site interaction of *E.coli* fRMsr and BPSL2418_{MES}. Hydrogen bonds are shown as green dashed lines between atoms. Hydrophobic interactions are indicated by a red lashes-like symbol pointing towards the ligand atom they contact. The MES atoms that are involved in hydrophobic interaction are represented by red spokes pointing back. (oxygen atom in red, nitrogen in blue, carbon in black, sulfur in yellow). This Figure was produced by LigPlot (Wallace *et al*, 1995).

6.1.1.2 BPSL2418_{Met-SO}

The crystal structure of BPSL2418 in complex with the substrate free Met-*R*-SO was solved at 1.4 Å. This structure belongs to the space group P2₁ and it consists of a dimer in the asymmetric unit with the substrate in bound to subunit. The overall structure and the active site of each BPSL2418_{Met-SO} subunit (Figure 6.7) is almost identical to BPSL2418_{MES} except for a slight change in loop 4 in the active site; the comparison between the two structures will be discussed later on in this chapter.

The Active Site: Each active site of the BPSL2418_{Met-SO} dimer is composed of the same BPSL2418_{MES} active site structure of four β-strands (β1, β2, β5 and β6) that are located in the center of the subunit and the (loop1- β3- loop2- α3- loop3- β4- loop4) construct that shields the binding ligand forming a small cavity. This cavity is made more tightened by the formation of the disulfide bond (CYS75-CYS109) that binds loop1 with loop4. The formation of this cavity in BPSL2418_{MES} and BPSL2418_{Met-SO} might ensure tight binding for the ligand. One molecule of free Met-*R*-SO substrate (Figure 6.8a) is bound within each active sites of the BPSL2418_{Met-SO} dimer identically. The substrate lies along the full-length of the central β-strands and the carboxylate group is located close to CYS85 in α3 (Figure 6.8b). The orientation of the free Met-*R*-SO carboxylate toward the free cysteine in α3 is also seen in the *N. meningitides* fRMsr (PDB code 3MMH) (Gruez *et al*, 2010) and the *S. pneumoniae* fRMsr (PDB code 3RFB) (Bong & Chi, 2011) structures. LigPlot (Figure 6.9) was used to investigate the interaction of the BPSL2418 protein and the free Met-*R*-SO substrate. These analyses indicate the residues in the active site that interact with the ligand: CYS109, ILE107 and HIS106 in loop4, VAL84 and CYS85 in α3, ILE78 and CYS75 in loop1, TRP53 and TYR57 in β1, ASP133 and ASP135 in β5, GLU116 in β4 and GLY83 in loop2. Hydrogen bonds occur between the amino nitrogen atom of Met-*R*-SO substrate and the OE1 and O atoms of GLU116 and ILE107 respectively (Figure 6.9). Also between the peptide nitrogens of ILE107 and VAL84 and the oxygen atom O of the Met-*R*-SO carboxylate group (Figure 6.9). Another hydrogen bond was predicted between the OXT of the Met-*R*-SO carboxylate and the peptide nitrogen of CYS85 in α3 (Figure 6.9). Other residues in the active site provide hydrophobic interactions with the free Met-*R*-SO, which ensures the

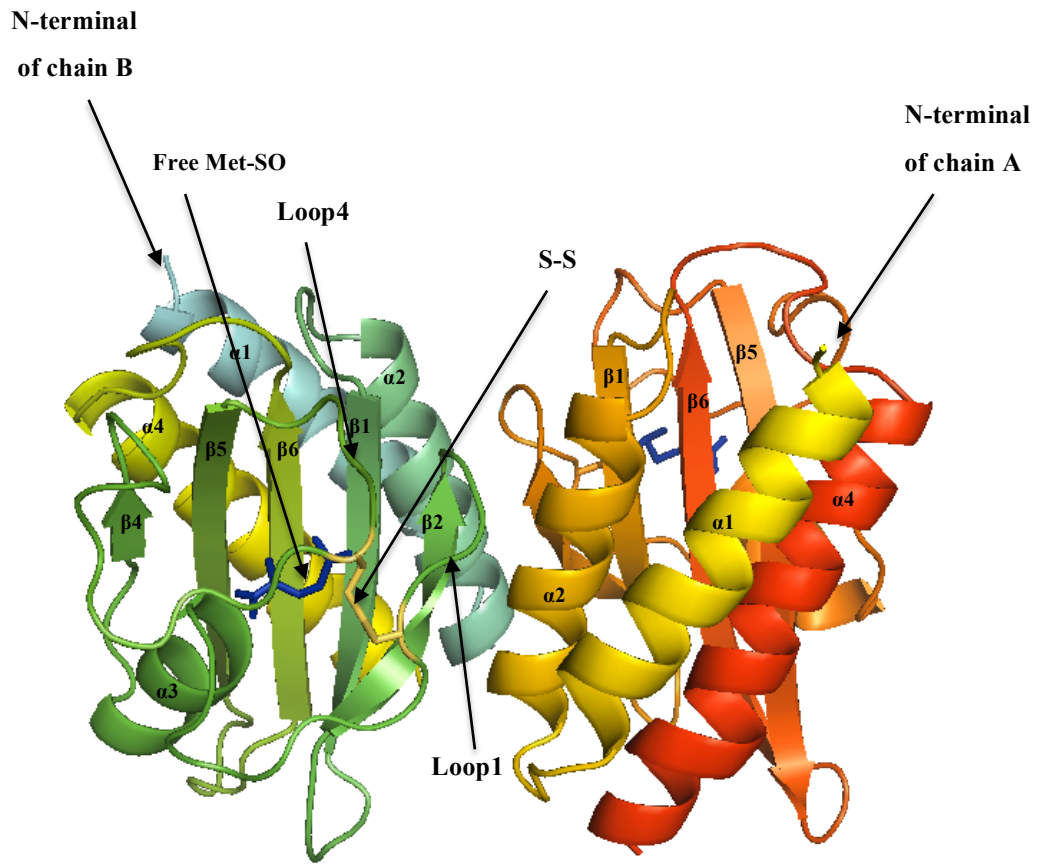


Figure 6.7: Cartoon representation of the overall fold structure of BPSL2418_{Met-SO}. The structure consists of a dimer with the free Met-SO (blue) substrate bound in each subunit of the asymmetric unit (on the right side chain A of BPSL2418_{Met-SO} dimer and on the left side chain B). Each subunit is composed of six β -strands and four α -helices with connecting loops. The disulfide S-S bond between CYS75 and CYS109 is shown in yellow color and free Met-SO molecules in blue color. The Figure was created by PyMol (DeLano & Lam, 2005).

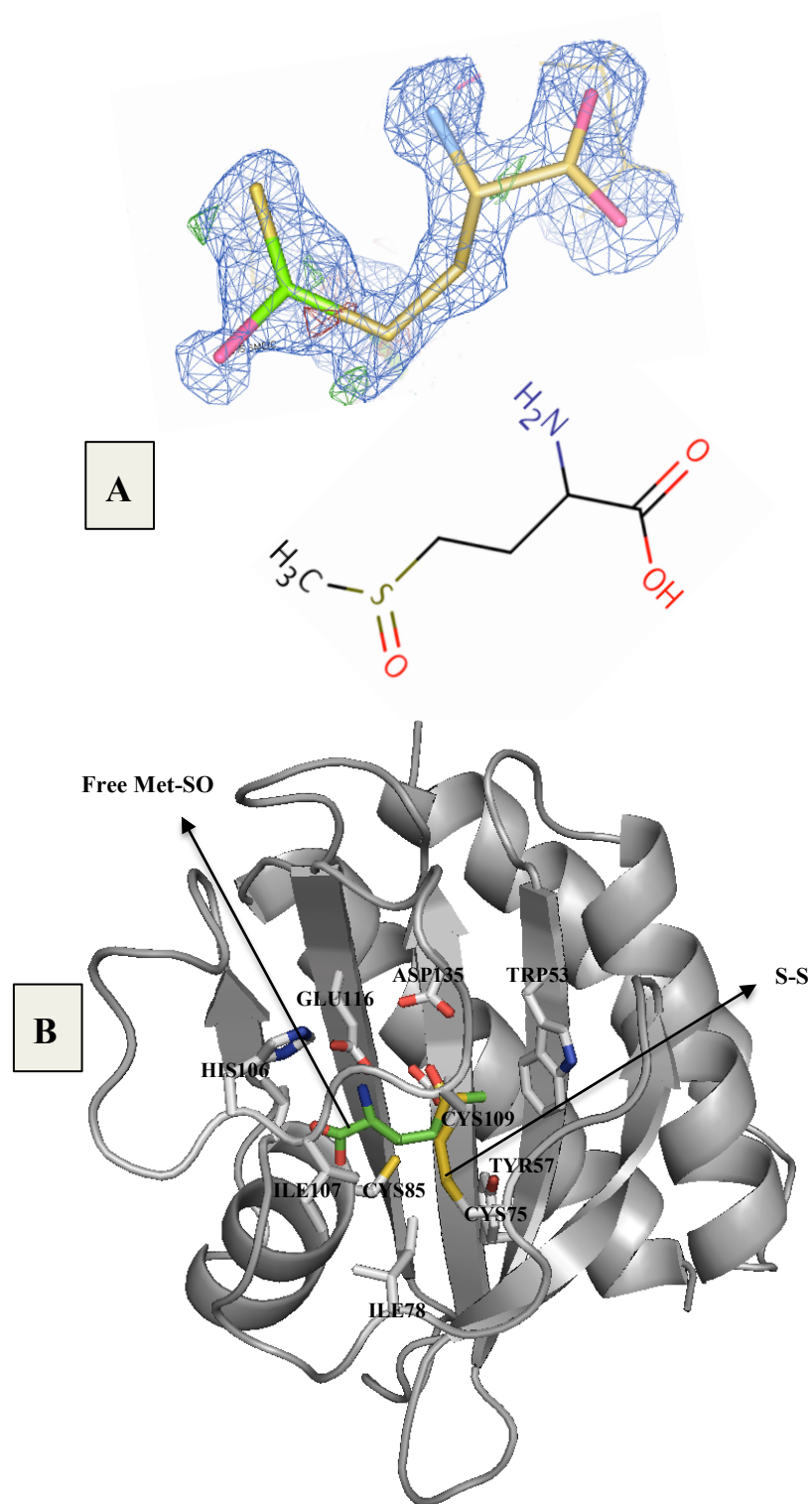


Figure 6.8: (A) Electron density map of free L-Methionine sulfoxide [free Met-*R*-SO] structure at 1.4 Å. (oxygen atom in red, nitrogen in blue and sulfur in green) (B) The binding site of BPSL2418Met-SO with free Met-SO substrate bound. Residues that are involved in BPSL2418-Met-SO binding are shown as sticks, the Figures were created by PyMol (DeLano & Lam, 2005). Blue map calculated using coefficients $2mF_o - dF_c$ and contoured at 0.4σ .

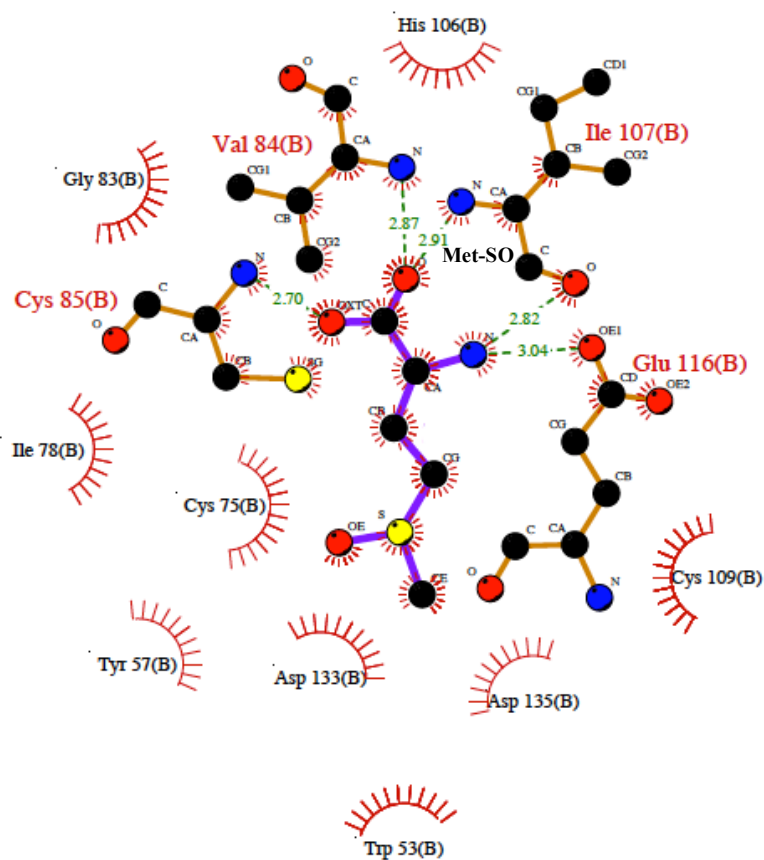


Figure 6.9: The binding site interaction of BPSL2418_{Met-SO} with the free Met-SO substrate. The Figure shows the residues, which involved in hydrogen bonds and hydrophobic interactions. Hydrogen bonds are shown as green dashed lines between atoms. Hydrophobic interactions are indicated by a red lashes-like symbol pointing towards the ligand atom they contact. The Met-SO atoms that involved in hydrophobic interaction are represented by a red spokes pointing back. (oxygen atom in red, nitrogen in blue, carbon in black, sulfur in yellow). This Figure was produced by LigPlot (Wallace *et al*, 1995).

tight binding of the substrate (Figure 6.9).

Similar 3D structures: Several homologs with high Z-scores and sequence identity have been predicted using the Dali server. As expected the Dali search revealed that BPSL2418_{Met-SO} belongs to the GAF domain fRMsR family (table 6.2). The first two hits are the *S. pneumonia* fRMsR (PDB code 3RFB) (Bong & Chi, 2011) and the *N. meningitides* fRMsR (PDB code 3MMH) (Gruez *et al*, 2010). The search gave a superposition of BPSL2418_{Met-SO} with *S. pneumonia* fRMsR with a Z-score of 27.3, sequence identity 42% and a root mean-square deviation (RMSD) of 1.2 Å between 153 α-carbon atom pairs (Figure 6.10). The superposition of BPSL2418_{Met-SO} was with *N. meningitides* fRMsR and gave a Z-score of 27.2, sequence identity 46% and a root mean-square deviation (RMSD) of 0.9 Å between 152 α-carbon atom pairs (Figure 6.11). The overall structure of BPSL2418_{Met-SO}, *S. pneumonia* fRMsR and *N. meningitides* fRMsR subunits are almost identical and share the GAF domain topology of four six antiparallel β-strands and four α-helices with connecting loops. An additional small alpha helix (ASP99, VAL100, ASP101, ALA102 and HIS103) in loop4 in *N. meningitides* fRMsR was the only difference between the three overall structures (Figure 6.10). The three structures have the three cysteines, which are conserved in all fRMsR family and a disulfide bond in the active site pocket. Like BPSL2418_{Met-SO}, each of the two homologs' structures consists of a dimer with the Met-SO substrate in each subunit of the asymmetric unit. The active sites of *S. pneumonia* fRMsR and *N. meningitides* fRMsR have a similar set of residues to the BPSL2418_{Met-SO} active site residues that provides hydrogen bonds and hydrophobic surface (Figure 6.10 & 6.12). The Met-R-SO lies inside the active sites of BPSL2418_{Met-SO}, *S. pneumonia* fRMsR and *N. meningitides* fRMsR identically, where the carboxylate of the substrate is close to the free cysteine in α3 (Figure 6.11). It has been mentioned in section 1.2.3 (2)&(3) the difference of the orientation of the position of Met-R-SO between the substrate bound form of *Staphylococcus aureus* fRMsR and the *Neisseria meningitides* fRMsR, where the sulfoxide moiety in the *Staphylococcus aureus* fRMsR replaces the position of the carboxylate group in *Neisseria meningitides* fRMsR and vice versa. This thesis shows that BPSL2418_{Met-SO} has the same orientation of the position of Met-R-SO of the *Neisseria meningitides* fRMsR,

Number	PDB-Chain	Z-score	R.M.S.D	Aligned residues	Length of PDB model	Sequence identity (%)	Molecule description
1	3 rfb-A	27.3	1.2	153	161	42	Structure of fRMsR of <i>Streptococcus pneumonia</i> (To be published) Method: X-ray diffraction Resolution: 2.3 Å Ligand: Met-R-SO
2	3 mmh-B	27.2	0.9	152	168	46	Structure of fRMsR of <i>neisseria meningitides</i> (Gruez et al, 2010). Method: X-ray diffraction Resolution: 1.25 Å Ligand: Met-R-SO
3	3 mmh-A	27.1	1.0	152	168	46	Structure of fRMsR of <i>neisseria meningitides</i> (Gruez et al, 2010). Method: X-ray diffraction Resolution: 1.25 Å Ligand: Met-R-SO
4	1 vhm-A	26.8	1.1	152	159	46	Structure of fRMsR of <i>E. coli</i> (Badger et al, 2005) Method: X-ray diffraction Resolution: 2.1 Å Ligand: MES
4	3 rfb-B	26.7	1.3	153	164	42	Structure of fRMsR of <i>Streptococcus pneumonia</i> (To be published) Method: X-ray diffraction Resolution: 2.3 Å Ligand: -

Table 6.2: Dali server results for the model of BPSL2418_{Met-so}. The top five hits are listed alongside their related Zscores, RMSD scores, alignment statistics and a brief description of each protein.

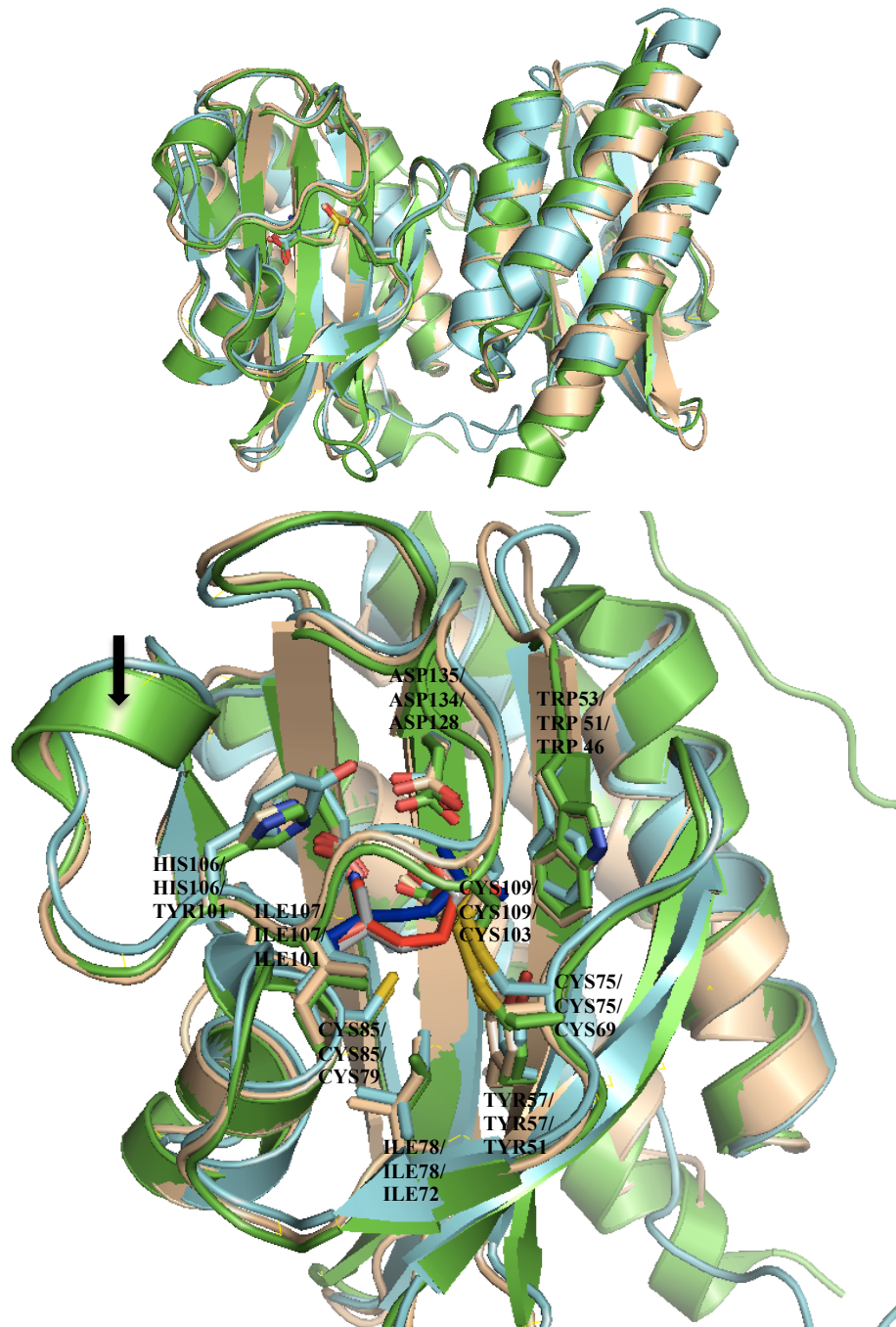


Figure 6.10: Superimposition of BPSL2418_{Met-SO} (gold), *N. meningitides* fRMsr (PDB code 3MMH) (green) and *S. pneumonia* fRMsr (PDB code 3RFB) (blue). (A) Overall fold of BPSL2418_{Met-SO}, 3MMH and 3RFB dimers. (B) The active sites of the three structures showing the binding residues and Met-SO ligand as sticks. The names of the corresponding residues are ordered as BPSL2418_{Met-SO} then 3MMH then 3RFB. Met-SO substrates are shown in red, grey and navy respectively. It is clear that the overall structures and the active sites of the three structures are almost identical except the small alpha helix in *N. meningitides* fRMsr pointed by black arrow. The Figure produced by PyMol (DeLano & Lam, 2005).

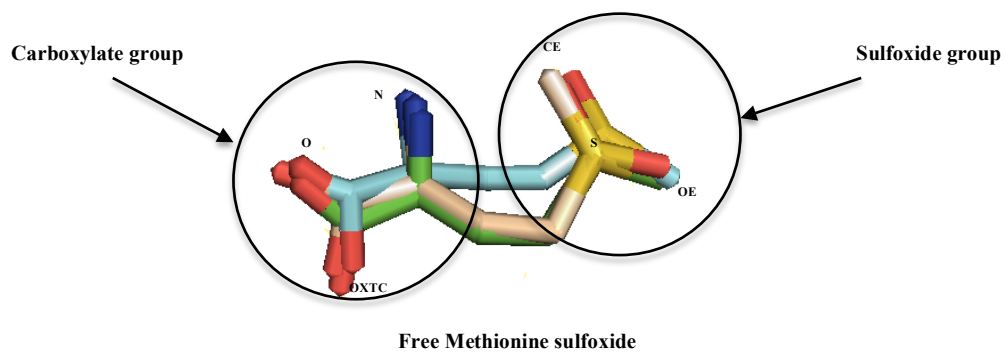
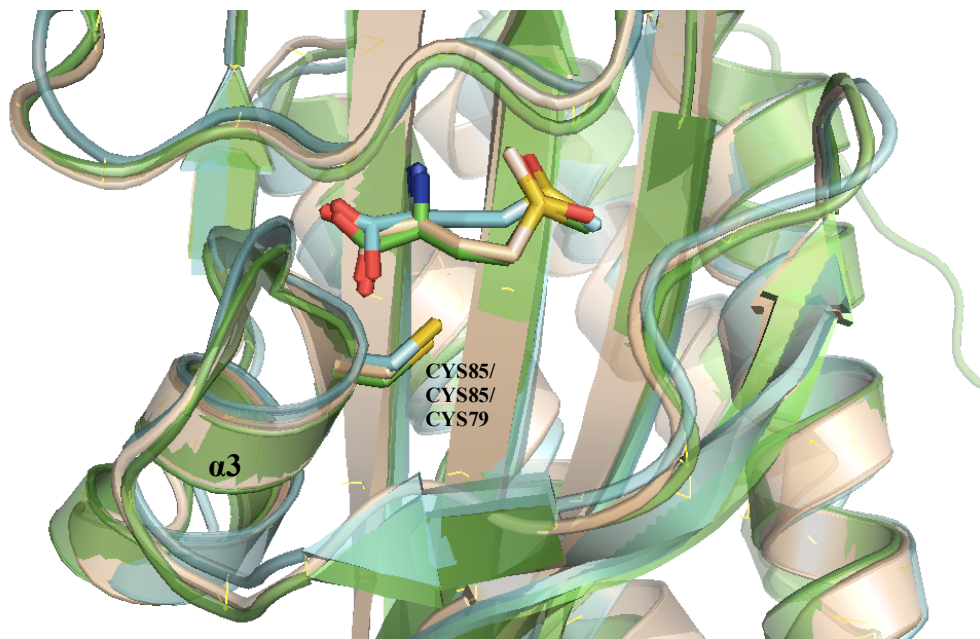


Figure 6.11: Close view of the Met-SO substrate that lies inside the active sites of BPSL2418_{Met-SO} (gold), *N. meningitides* fRMsR (PDB code 3MMH) (green) and *S. pneumoniae* fRMsR (PDB code 3RFB) (blue). The Figure shows that the carboxylate groups of the substrate inside the three active sites are located close to the free cysteine in α_3 . The Figure produced by PyMol (DeLano & Lam, 2005).

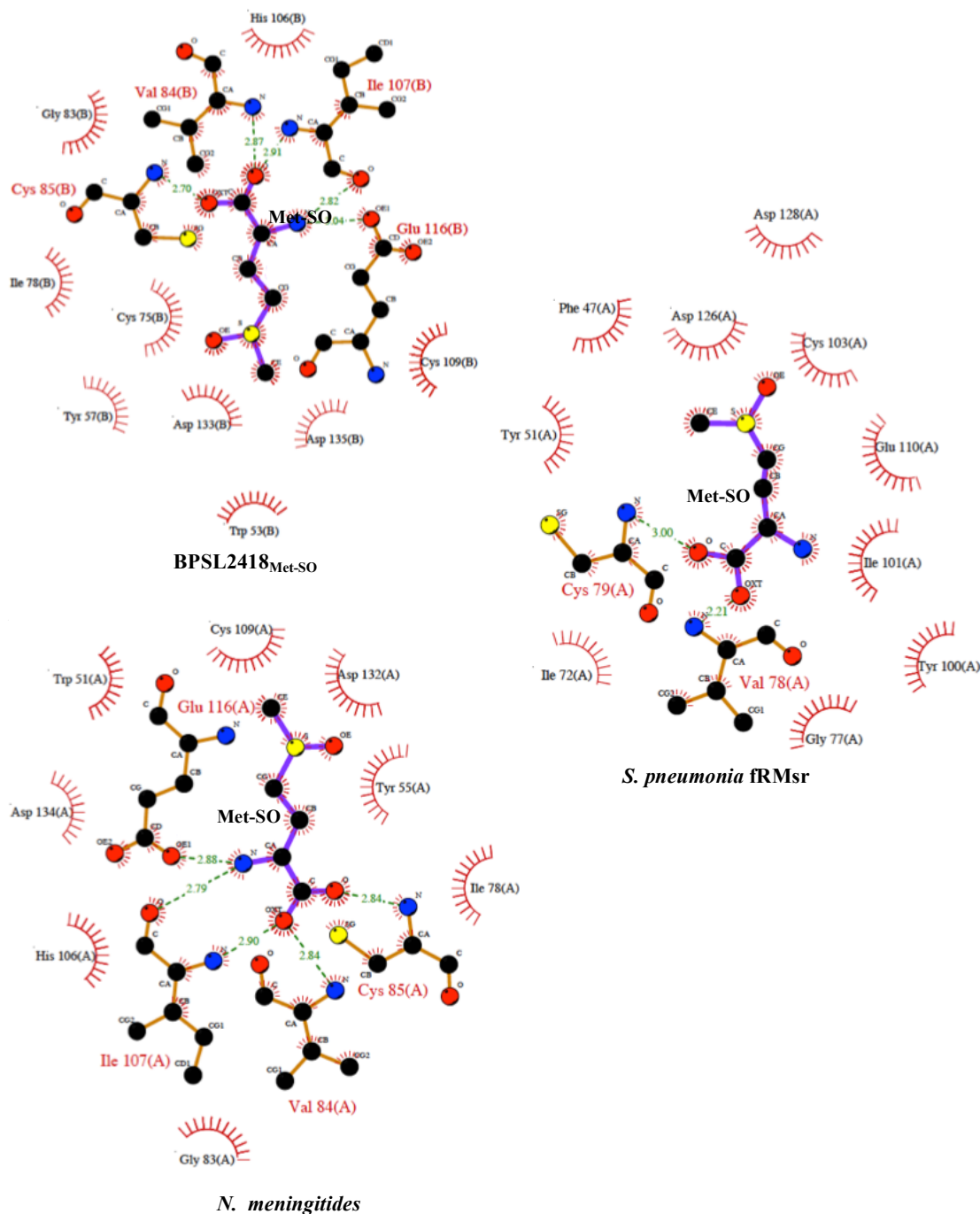


Figure 6.12: LigPlot diagrams of the binding site interaction of BPSL2418_{Met-SO}, *N. meningitidis* fRMs and *S. pneumoniae* fRMs. Hydrogen bonds are shown as green dashed lines between atoms. Hydrophobic interactions are indicated by a red lashed-like symbol pointing towards the ligand atom they contact. The Met-SO atoms that involved in hydrophobic interaction are represented by a red spokes pointing back. (oxygen atom in red, nitrogen in blue, carbon in black, sulfur in yellow). This Figure was produced by LigPlot (Wallace *et al*, 1995).

S. pneumoniae fRMs_r and *N. meningitidis* fRMs_r and only the *Staphylococcus aureus* fRMs_r has different orientation. It's not clear why the *Staphylococcus aureus* fRMs_r structure binds the substrate in different orientation, but this might be because it was obtained using C68S mutant (CYS68 corresponding to CYS75 in BPSL2418). This feature is different from the other fRMs_r-substrate complex are oxidized forms with a disulfide bond between two cysteines in the active site.

6.1.1.3 BPSL2418_{Reduced}

In order to investigate changes in the active site of the BPSL2418 protein that predicted to act as free methionine sulfoxide reductase, a reduced form of BPSL2418 was produced. The BPSL2418_{Reduced} crystal, which diffracted to 2.0 Å, comprised two subunits in the asymmetric unit. The structure of the BPSL2418 protein has changed in the active site of the different forms. A comparison between the reduced BPSL2418 and the other two oxidized forms BPSL2418_{MES} and BPSL2418_{Met-SO} will be illustrated in the following sections.

The Active Site: The overall structure of one subunit involved six β strands and four α helices (Figure 6.13). The active site as in the other forms of BPSL2418 is composed of the four β-strands (β1, β2, β5 and β6) bounded by (α1, α2 and α4) helices from one side and the (loop1- β3- loop2- α3- loop3- β4- loop4) unit from the other side. The reduced form has no bound ligand in either active sites, and has no disulfide bond between CYS75-CYS109 due to the addition of the DTT reductant in the crystallization solution (Figures 5.22 & 6.14).

Similar 3D Structures: Like the other two forms of BPSL2418, a search with Dali server with BPSL2418_{Reduced} shows several homologs belonging to fRMs_r family (table 6.3) with a GAF domain topology. The search gave a best superimposition of BPSL2418_{Reduced} with the reduced form of *S. aureus* fRMs_r (PDB code 3KSG) (Bong *et al*, 2010) with a Z-score of 26.5, sequence identity 44% and a root mean-square deviation (RMSD) of 1.0 Å between 151 α-carbon atom pairs (Figure 6.15). The superimposition reveals that the BPSL2418_{Reduced} and *S. aureus* fRMs_r share a similar fold except the present of an extra small helix between β4 and β5 in *S. aureus* fRMs_r.

Number	PDB-Chain	Z-score	R.M.S.D	Aligned residues	Length of PDB model	Sequence identity (%)	Molecule description
1	3 mmh-B	26.5	1.0	151	168	44	Structure of fRMs _r of <i>neisseria meningitides</i> (Gruez et al, 2010). Method: X-ray diffraction Resolution: 1.25 Å Ligand: Met-R-SO
2	3 mmh-A	26.4	1.0	151	168	44	Structure of fRMs _r of <i>neisseria meningitides</i> (Gruez et al, 2010). Method: X-ray diffraction Resolution: 1.25 Å Ligand: Met-R-SO
3	3 rfb-A	26.3	1.5	151	161	40	Structure of fRMs _r of <i>Streptococcus pneumonia</i> (To be published) Method: X-ray diffraction Resolution: 2.3 Å Ligand: Met-R-SO
4	1vhm-A	26.3	1.1	151	159	48	Structure of fRMs _r of <i>E. coli</i> (Badger et al, 2005) Method: X-ray diffraction Resolution: 2.1 Å Ligand: MES
5	1vhm-B	26.0	1.2	151	160	48	Structure of fRMs _r of <i>E. coli</i> (Badger et al, 2005) Method: X-ray diffraction Resolution: 2.1 Å Ligand: MES

Table 6.3: Dali server results for the model of BPSL2418_{Reduced}. The top five hits are listed alongside their related Zscores, RMSD scores, alignment statistics and a brief description of each protein.

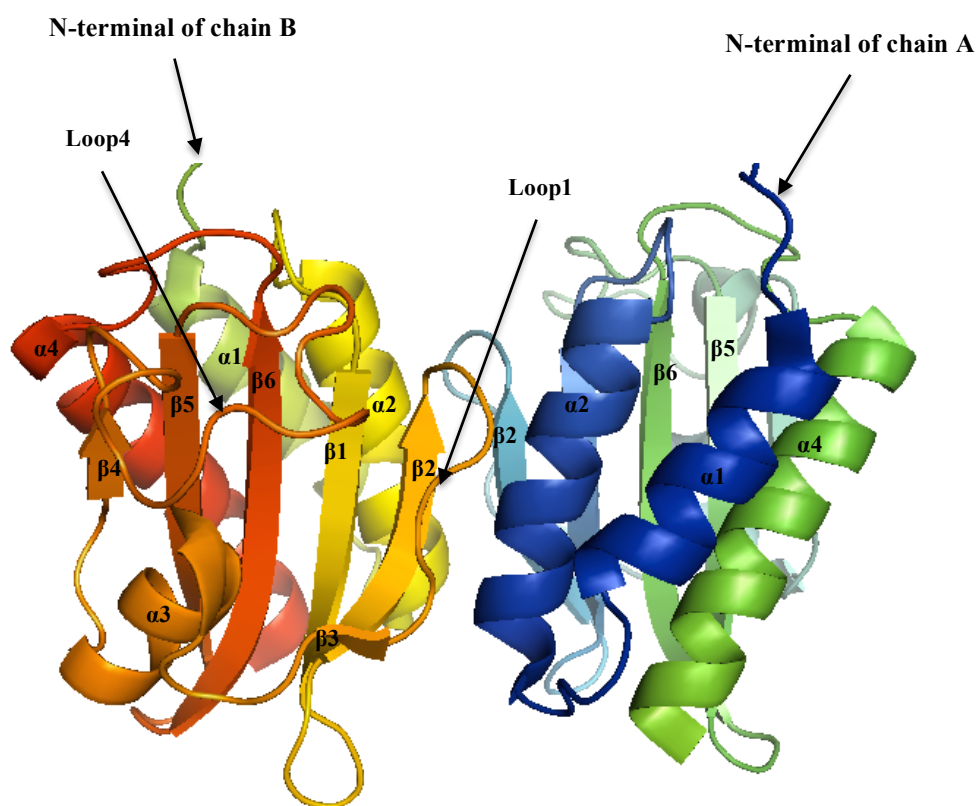


Figure 6.13: Cartoon representation of the overall fold structure of BPSL2418_{Reduced}. The structure consists of a dimer with no disulfide bond formed in any of each subunit (on the right side chain A of BPSL2418_{Reduced} dimer and on the left side chain B). Each subunit is composed of six β-strands and four α-helices with connecting loops. The Figure was created by PyMol (DeLano & Lam, 2005).

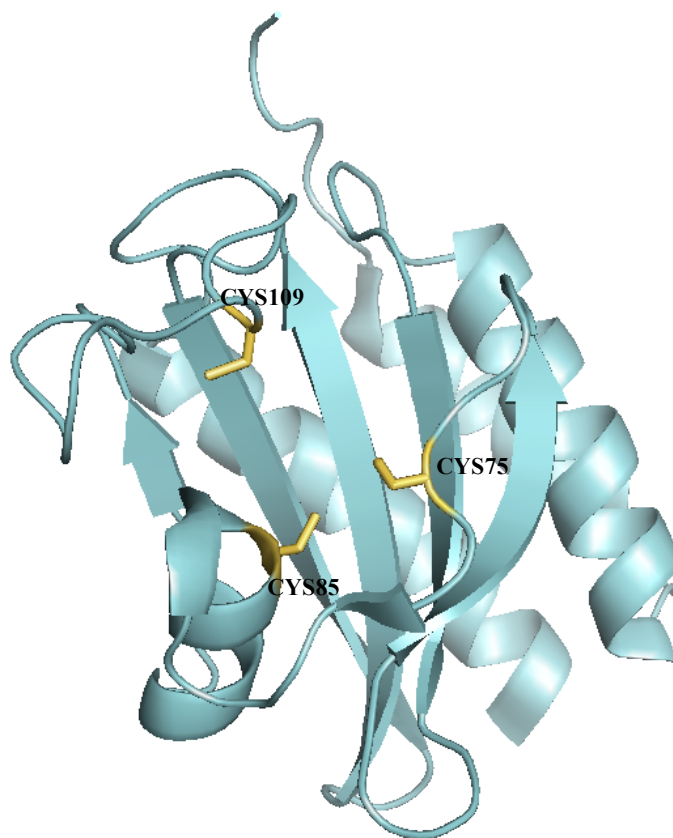


Figure 6.14: Cartoon representation of BPSL2418_{Reduced} structure showing the active site. The active site contains the three conserved cysteines of the fRMs_r family as sticks. No disulfide bond was formed and no bound ligand is in the active site.

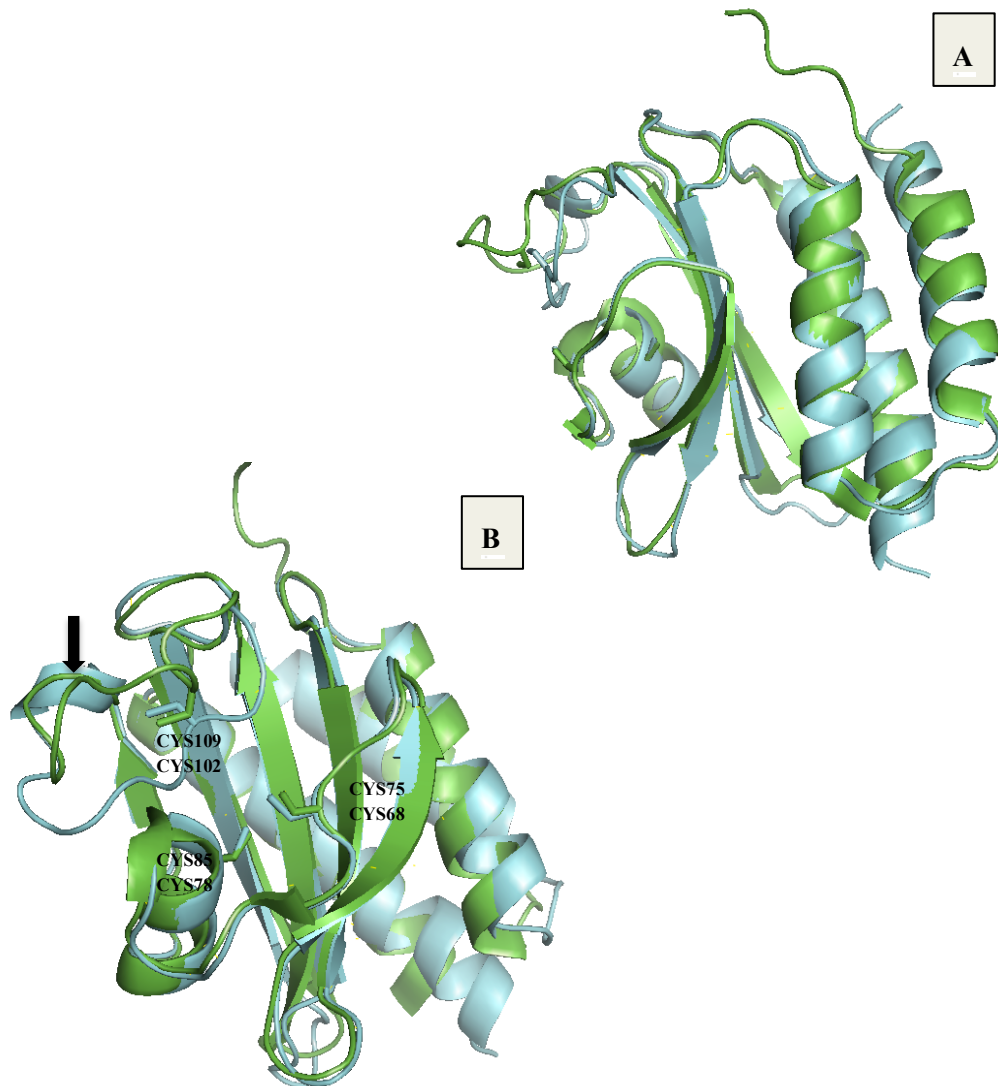


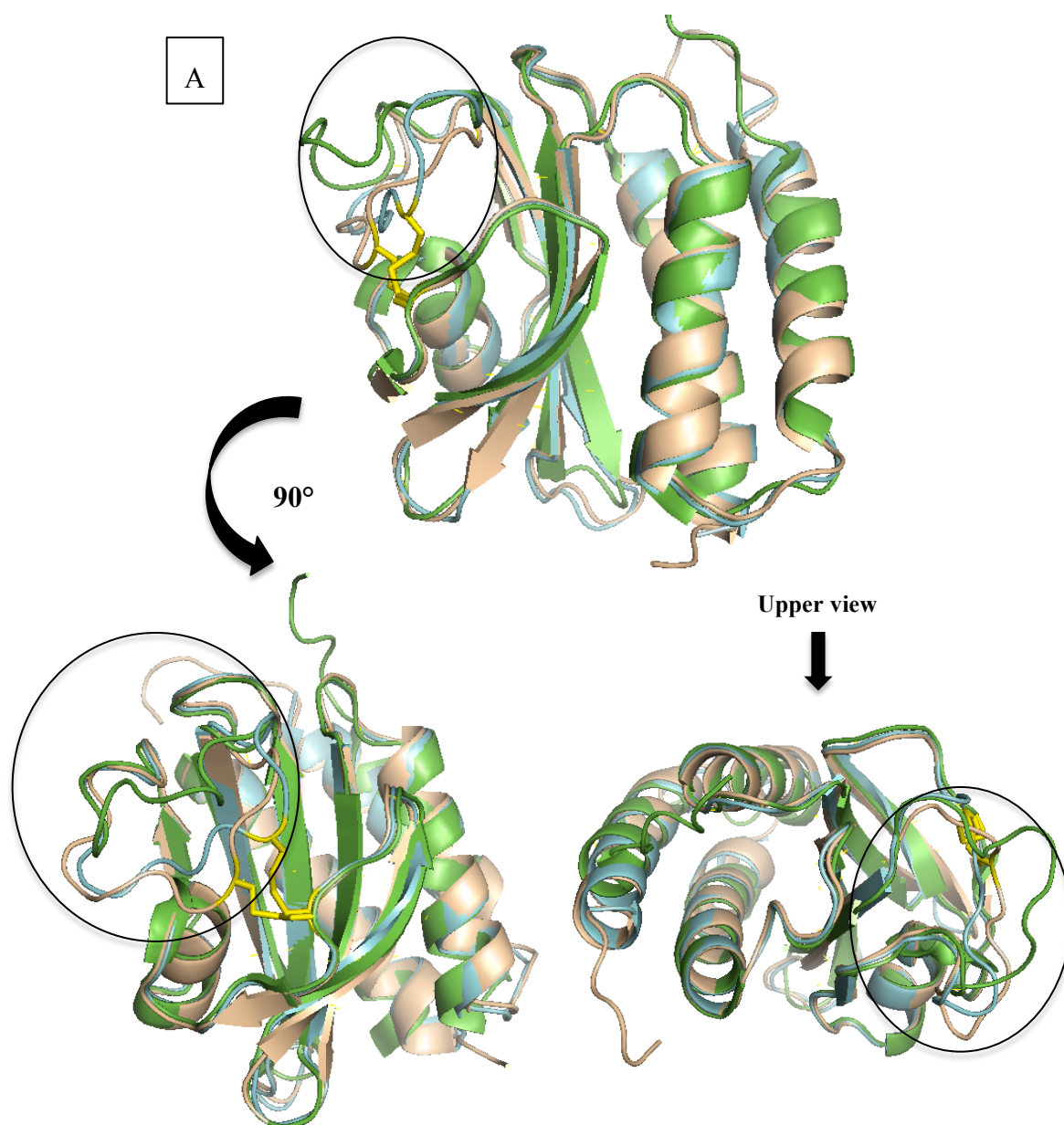
Figure 6.15: Superimposition of BPSL2418_{Reduced} (green) chain A and *S. aureus* fRMsR (PDB code 3KSG) (light blue) chain A. (A) Side view, (B) active site of the BPSL2418_{Reduced} and *S. aureus* fRMsR. The superimposition shows that the overall structures and the active sites of the two structures are matching well except the extra alpha helix between $\beta 4$ and $\beta 5$ in *S. aureus* fRMsR (pointed by a black arrow). The three cysteines residues in the active site of the two proteins are shown as sticks. The Figure produced by PyMol (DeLano & Lam, 2005).

6.1.2 Comparison and conformational changes of MES-bound, Met-SO-bound and reduced forms of BPSL2418

In this section a comparison between the BPSL2418_{MES}, BPSL2418_{Met-SO} and BPSL2418_{Reduced} structures, which are expected to represent the catalytic steps of the BPSL2418 reaction will be described.

6.1.2.1 Comparison of the overall structures of BPSL2418 forms

The overall structural alignment of BPSL2418_{MES}, BPSL2418_{Met-SO} and BPSL2418_{Reduced} was produced using PyMol to compare the three forms of BPSL2418 (Figure 6.16). The alignment shows significant conformational changes between the three forms in the active site. No change was observed in loop1 of BPSL2418_{MES}, BPSL2418_{Met-SO} and BPSL2418_{Reduced} and the position of CYS75 residues of the three forms was unchanged (Figure 6.17). A conformational change was shown in loop4, which is composed of residues 98-113 including CYS109 (Figure 6.17). It is clear that the CYS109 residue of each loop4 of the three structures is located at a different position. The C α of the CYS109 residue of BPSL2418_{MES}, BPSL2418_{Met-SO} and BPSL2418_{Reduced} resides at a distance of 6.0 Å, 5.7 Å and 10.0 Å, respectively from the C α of the CYS75 residue in loop1 of each form. This indicates that the loop4 of BPSL2418_{Met-SO} moves toward the entrance of the active site, where a disulfide bond is produced between CYS109 and CYS75, leading to closure of the active site pocket. The BPSL2418_{MES} structure is an oxidized form and the distance between the C α of CYS109 and CYS75 is slightly longer compared with the C α - C α distance between the two cysteines in BPSL2418_{Met-SO}. Also the CYS109 residue in BPSL2418_{MES} pulls loop4 to the entrance of the active site producing a cavity, but it seems to be less tight than the BPSL2418_{Met-SO} pocket (Figure 6.17). A large movement far from the entrance of the cavity occurs in loop4 of the reduced form. The C α of CYS109 in loop4 of BPSL2418_{Met-SO} and BPSL2418_{Reduced} reside at a distance of 4.6 Å from each other (Figure 6.18). Also the sulfur atom of CYS109 in the reduced form points toward the opposite side of the active site cavity (Figure 6.17). The volume of the active site pocket of the three forms was calculated using CASTp <http://sts-fw.bioengr.uic.edu/castp/> BPSL2418_{Reduced} has the largest pocket volume of



B

Structure	Resolution	RMSD
BPSL2418_{MES}	1.18 Å	
Chain A		0
Chain B		0
BPSL2418_{Met-SO}	1.4 Å	
Chain A		0.37 Å
Chain B		0.35 Å
BPSL2418_{Reduced}	2.0 Å	
Chain A		0.39 Å
Chain B		0.37 Å

Figure 6.16: Different views of the superimposition of BPSL2418_{MES} (gold), BPSL2418_{Met-SO} chain A (blue) and BPSL2418_{Reduced} chain A (green) with disulfide bonds in yellow. (A) The alignment shows an overall very good superimposition but with a significant change in loop4 (defined by black circle). (B) RMSD between subunits after superimposition. The Figure produced by PyMol (DeLano & Lam, 2005).

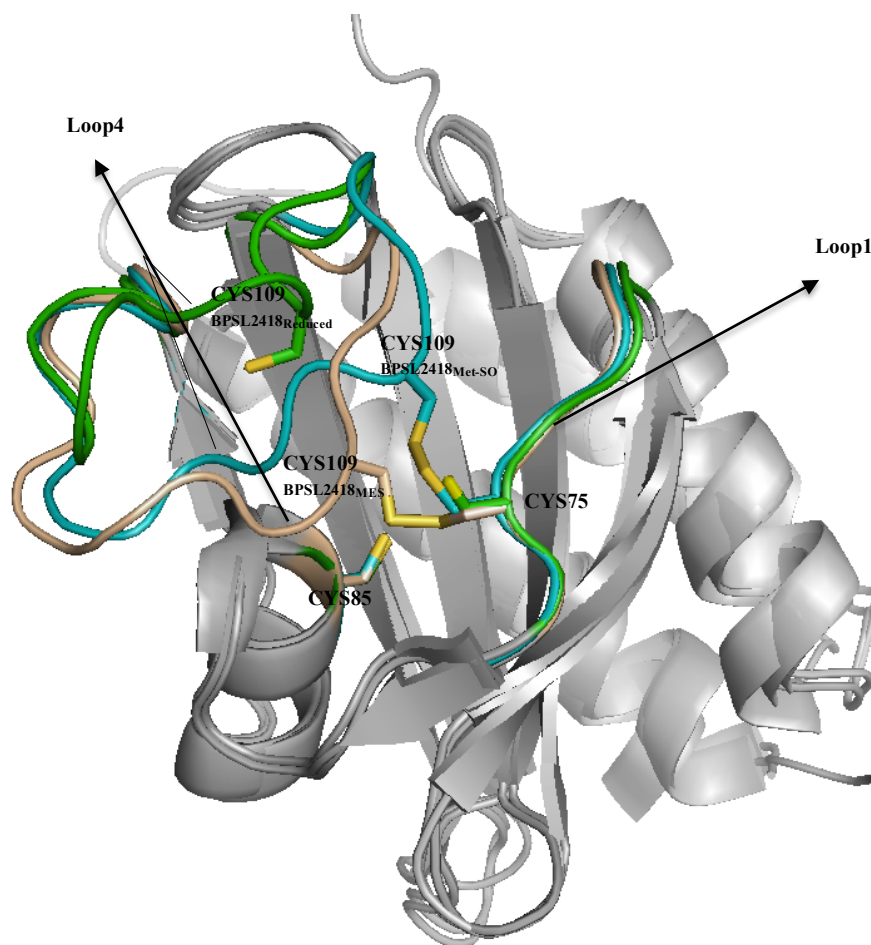


Figure 6.17: Superimposition of the three forms of BPSL2418 shows the change in the active site, BPSL2418_{MES} (gold), BPSL2418_{Met-SO} (blue) and BPSL2418_{Reduced} (green). The Figure shows the positions of the three conserved cysteines CYS75, CYS85 and CYS109 in the active sites. The cysteines side chains are shown as sticks. The Figure produced by PyMol (DeLano & Lam, 2005).

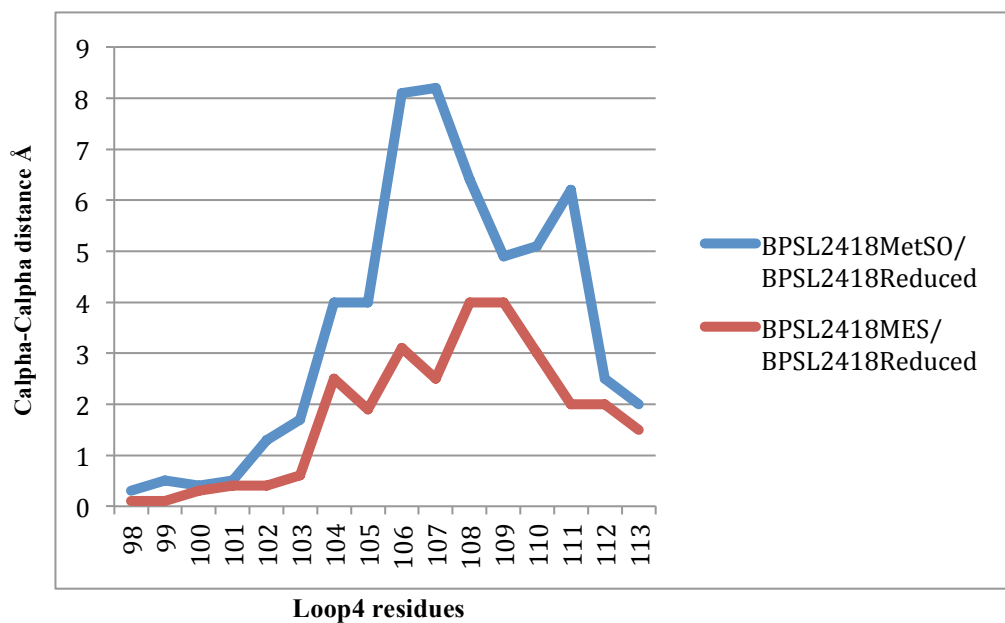


Figure 6.18: Calpha-Calpha distance (Y-axis) of residues of loop4 between BPSL2418_{Met-SO} and BPSL2418_{Reduced} (blue line) and between BPSL2418_{MES} and BPSL2418_{Reduced} plotted against residues number of loop4 (X-axis).

445.5 Å³, the BPSL2418_{MES} has a pocket volume of 204.2 Å³ and the BPSL2418_{Met-SO} has the smallest pocket volume of 167.3 Å³. The B-value (temperature-factor) for the atoms of the loop4 compared to the rest of the structure in each form was calculated. Atoms with low B-factors indicate a well-ordered part of the structure. Atoms with high B-factors belong to a part of the structure that has a higher degree of mobility and flexibility. The B-factors of loop4 of subunit A of BPSL2418_{MES}, BPSL2418_{Met-SO} and BPSL2418_{Reduced} are (14.27, 19.5 and 32.8 Å² respectively), which are significantly higher than the overall B-factors of the subunit A of each form (8.52, 14.6 and 27.2 Å² respectively). It has been shown also that the B-factor of the main chain residues of BPSL2418_{Reduced} is higher than the main chain residues of BPSL2418_{MES} and BPSL2418_{Met-SO} (Figure 6.19). In sum the movement of the loop4, particularly CYS109, determines the conformational changes of the reduced form and the oxidized form (BPSL2418_{Met-SO} and BPSL2418_{MES}) by creating a closed conformation in BPSL2418_{Met-SO} and BPSL2418_{MES} and an open conformation in BPSL2418_{Reduced}. This result is consistent with the comparison between the reduced and oxidized forms of fRMSr from *S. aureus* (Bong *et al*, 2010). The conformational changes can also be observed by comparing the molecular surface of the three forms.

6.1.2.2 Comparison of the molecular surface of the different forms of BPSL2418

The electrostatic surfaces of the three different BPSL2418 forms were calculated using Pymol (DeLano & Lam, 2005). The two subunits for each BPSL2418_{Met-SO} and BPSL2418_{Reduced} dimers are identical, therefore only one subunit of each form will be used in the comparison. The models show the charge distribution on the surface of BPSL2418 forms (Figure 6.20). A fairly negatively charged depression appears in the entrance of the active site clearly shown in the reduced form (Figure 6.20 c). The negative potential in the active site possibly indicates the substrate-binding site. It is formed by the side chains of residues (ASP60, GLU63, ASP99, ASP110, GLU116, ASP125, ASP133, ASP135 and ASP143) in the active site. The negatively charged pocket might be significant in terms of the substrate binding. To investigate the role of these negatively charged residues, this can be examined

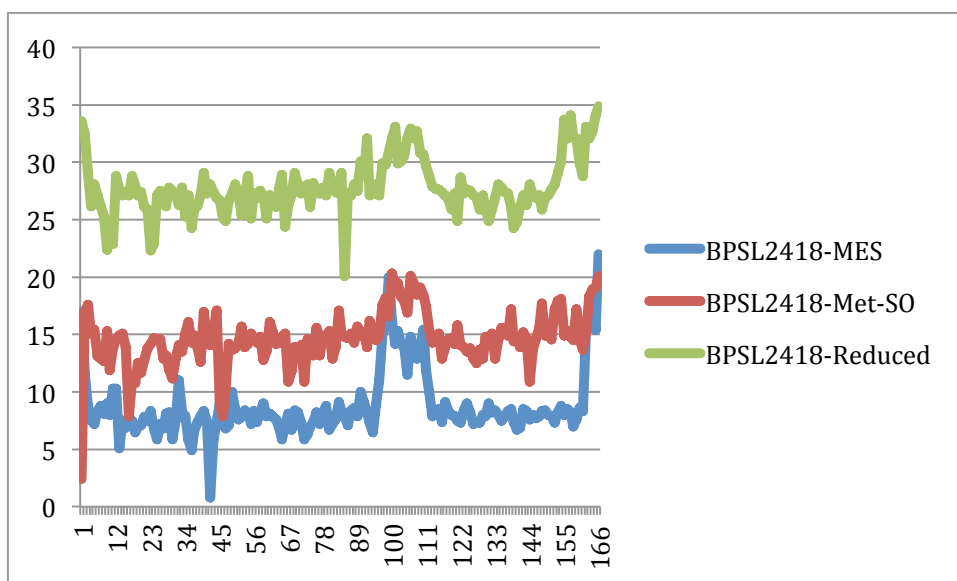


Figure 6.19: B-factors (\AA^2) of main chain residues of BPSL2418 different forms plotted against residues numbers (X-axis). The blue, red and green lines represent B-factors of main chain of BPSL2418_{MES}, BPSL2418_{Met-SO} and BPSL2418_{Reduced} respectively.

by site-directed mutagenesis, kinetic characterization and crystallographic analysis. In these Positive charge is distributed randomly on the surface of BPSL2418, provided by the side chains of LYS and ARG residues. The comparison of the molecular surface of BPSL2418 forms reveals that the reduced form creates an opened conformation (Figure 6.20 c) while the MES and the Met-SO complex forms produce a closed conformation of the active site cavity (Figure 6.20 b). However the molecular surface model shows that the BPSL2418_{Met-SO} form generates a more closed conformation and this might indicate that the free Met-R-SO is the most preferable ligand. The closed conformation status of the two oxidized forms (BPSL2418_{Met-SO} and BPSL2418_{MES}) supports the role of the disulfide bond in providing tight binding for the ligand in a small cavity that is covered by the flap (loop1- β 3- loop2- α 3- loop3- β 4- loop4 unit). In addition to offering tight binding to the substrate, this covered small cavity protects the active site from the solvent. The active site of many enzymes must protect their substrate from water because the reactions cannot proceed rapidly or even at all in an aqueous environment. The ability of some enzymes to shield their substrate from aqueous solvent by taking advantage of conformational changes that close off the active site from contact with bulk solvent is important in enabling them to accelerate the rates of the reactions they catalyze (Bernstein *et al* 1995).

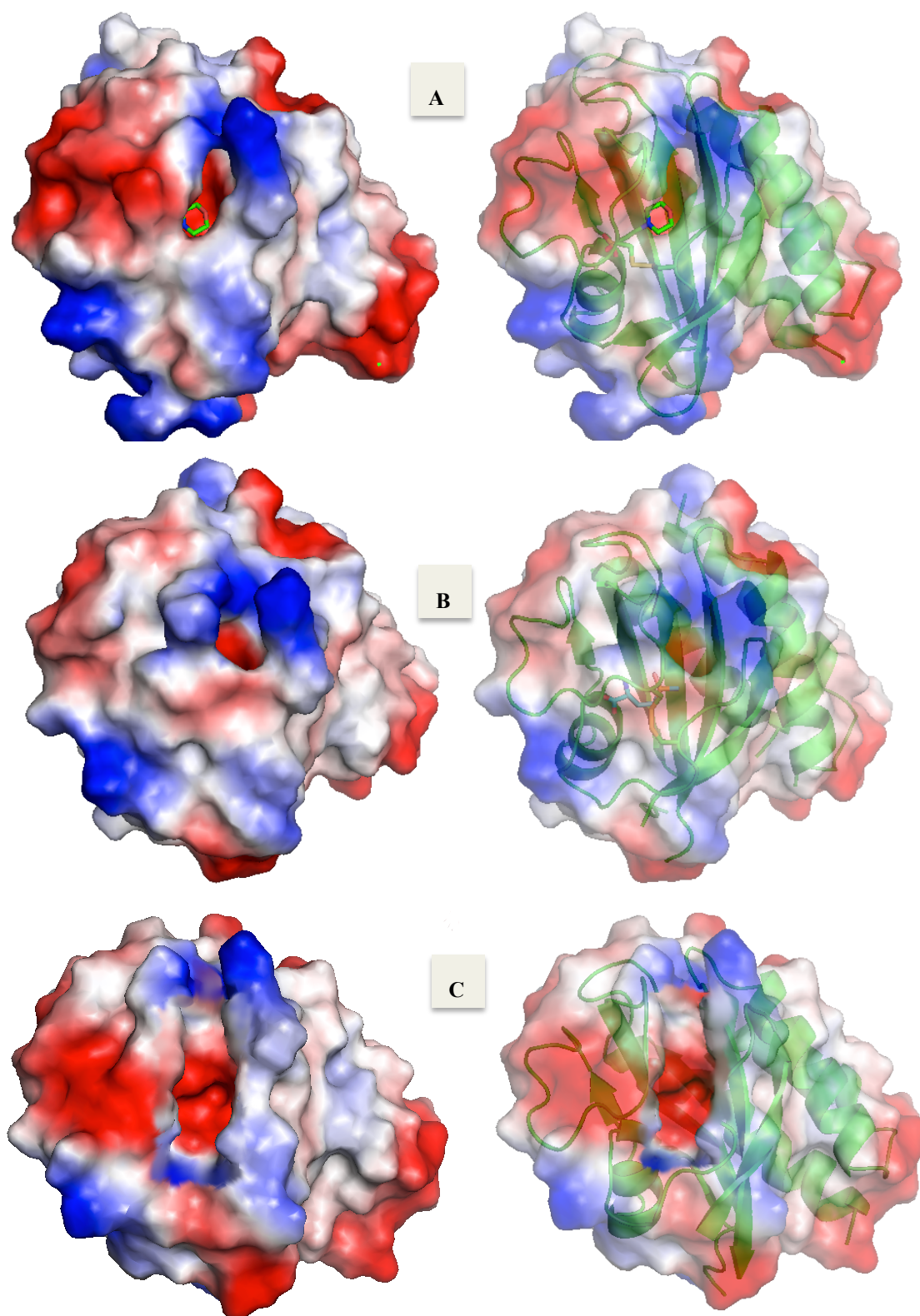


Figure 6.20: Comparison of electrostatic surface potential of BPSL2418 forms showing the conformational changes in the active site. (A) BPSL2418_{MES} creates a closed conformation. (B) BPSL2418_{Met-SO} creates a more closed conformation. (C) BPSL2418_{Reduced} creates an open conformation; the model shows the putative binding site inside the cavity. On the left side the subunits of the three forms are shown as solid electrostatic surface model, on the right side the subunit are shown as semi-transparent electrostatic surface and ribbon models. The Figures produced by PyMol (DeLano & Lam, 2005).

6.2 BPSL2418 interfaces, surfaces and assemblies

In protein crystallization, the molecules pack together in a regular array (Rhodes, 2006). In crystals, the interactions between protein molecules and their binding energy, interface area, hydrogen bonds, hydrophobic interactions, solvation energy and salt bridges across the various interfaces can be examined using PDBePISA webserver http://pdbe.org/PISA_ (Krissinel & Henrick, 2007). This is a web service for analyzing the macromolecular surfaces, interfaces, interactions and assemblies. The crystal packing analysis of the three forms of BPSL2418 using PDBePISA webserver shows that the subunits form dimers. This result corresponds with the apparent dimeric state indicated from the gel filtration profile of BPSL2418 (Figure 4.10).

6.2.1 BPSL2418_{MES} analysis

The BPSL2418_{MES} crystal contacts and molecular interactions, which maintain the molecules together in the crystal, were investigated using PDBePISA with a model containing the protein chain and the MES molecule.

6.2.1.1 BPSL2418_{MES} monomers

Chain A, binding MES ligand which represents the BPSL2418_{MES} protein structure was calculated using PISA to has a total of 162 residues, 146 of them are surface exposed residues. The solvent accessible area for chain A is 8000 Å² approximately and the solvation free energy gain from interface formation (ΔG) is estimated as -149.5 Kcal/mol.

6.2.1.2 Monomer-monomer interface forming dimeric BPSL2418_{MES}

The PISA analysis indicates that the most extensive crystal packing appears between BPSL2418_{MES} molecules related to each other by a 2-fold axis symmetry (Figure 6.21). PISA estimates a dimeric state for BPSL2418_{MES} where the complexation significance score CSS is equal 1. The CSS score shows how

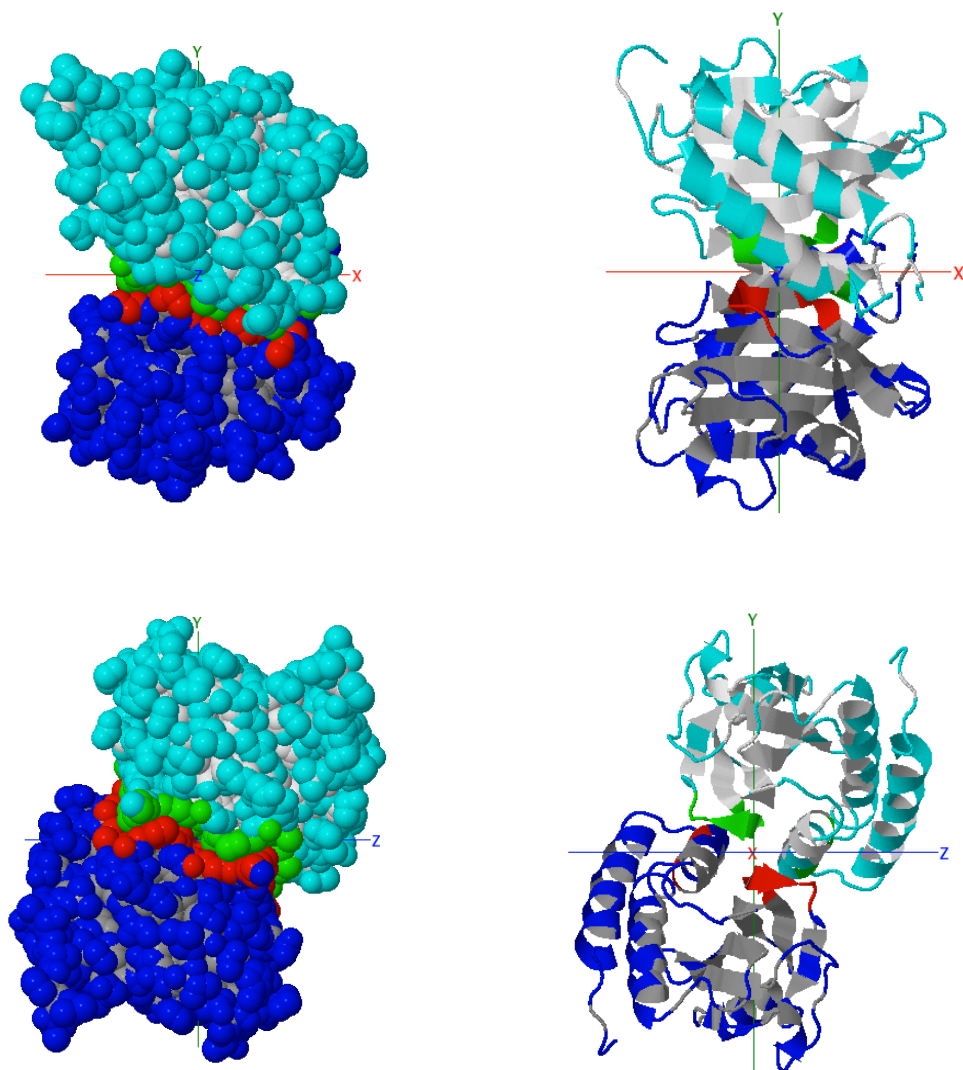


Figure 6.21: The dimeric structure model of BPSL2418_{MES}. On the left side the space-filling model and on the right side the cartoon model, each has three images taken 90° apart about the Y-axis. The two molecules forming the dimer are in blue ● and tiffany ● colors and the interface area in green ● and red ● colors. The Figure was produced using PISA webserver http://pdbe.org/PISA_

significant is the interface for assembly formation and it ranges from 0 to 1 as interface relevance to complex formation increases. The monomer-monomer interface forming dimeric BPSL2418_{MES} has a buried surface area of approximately 800 Å² including 27 residues from each molecule (table 6.4). The dimer interface involves 7 hydrogen bonds, 4 salt bridges (table 6.5) (Figures 6.22) and no disulfide bonds or covalent bonds. The estimated interface area represents 9.9% of the total surface area and lies in the normal buried surface area range of protein-protein interactions (5-25%) (Jones & Thornton, 1995). The PISA analysis is in a good agreement with the result from gel filtration (Figure 4.10) thus, it is most likely that BPSL2418_{MES} occurs as a dimer. The dimer interface is made up of residues in α_2 , α_1 - α_2 loop, β_2 and β_2 - β_3 loop interact with residues in α_2 , α_1 - α_2 loop, β_2 and β_2 - β_3 loop of the symmetry related molecule.

	Residue	Accessible surface area (Å²)	Buried surface area (Å²)
1	LEU 26	67.16	16.77
2	GLU 30	77.83	58.61
3	THR 31	59.31	3.68
4	ASP 32	43.94	21.45
5	VAL 34	92.70	57.89
6	ALA 35	39.77	39.77
7	ASN 36	2.38	1.84
8	ALA 38	42.51	40.18
9	ASN 39	71.65	70.77
10	SER 42	33.04	33.04
11	LEU 43	38.53	11.34
12	TYR 45	52.32	0.89
13	HIS 46	134.07	36.99
14	PHE 58	42.60	0.63
15	VAL 65	30.97	20.92
16	VAL 66	29.86	28.02
17	GLY 67	17.87	17.03
18	PRO 68	27.72	26.55
19	PHE 69	60.96	56.89
20	GLN 70	45.19	27.74
21	GLY 71	36.60	14.17
22	LYS 72	160.16	36.19
23	PRO 73	128.65	70.64
24	ALA 74	30.41	11.54
25	VAL 76	74.75	53.69
26	ARG 77	128.82	36.84
27	THR 166	69.20	1.35

Table 6.4: Accessibility and buried surface areas of 27 residues involved in the BPSL2418_{MES} monomer-monomer interface calculated by PISA <http://pdbe.org/PISA> (Krissinel & Henrick, 2007).

Hydrogen bonds

	Monomer I	Distance (Å)	Monomer II
1	ARG 77 [NE]	2.81	GLU 30 [OE1]
2	ARG 77 [NH2]	2.93	GLU 30 [OE2]
3	ALA 74 [N]	2.94	ASN 39 [OD1]
4	GLN 70 [NE2]	3.15	PHE 69 [O]
5	SER 42 [OG]	2.65	GLY 71 [O]
6	GLN 70 [NE2]	3.44	GLY 71 [O]
7	ASN 39 [ND2]	3.02	ALA 74 [O]

Salt bridges

	Monomer I	Distance (Å)	Monomer II
1	ARG 77 [NE]	2.81	GLU 30 [OE1]
2	ARG 77 [NH2]	3.72	GLU 30 [OE1]
3	ARG 77 [N]	3.55	GLU 30 [OE2]
4	ARG 77 [NE2]	2.93	GLU 30 [OE2]

Table 6.5: The hydrogen bonds and salt bridges formed in BPSL2418_{MES} monomer-monomer interface calculated by PISA <http://pdbe.org/PISA> (Krissinel & Henrick, 2007).

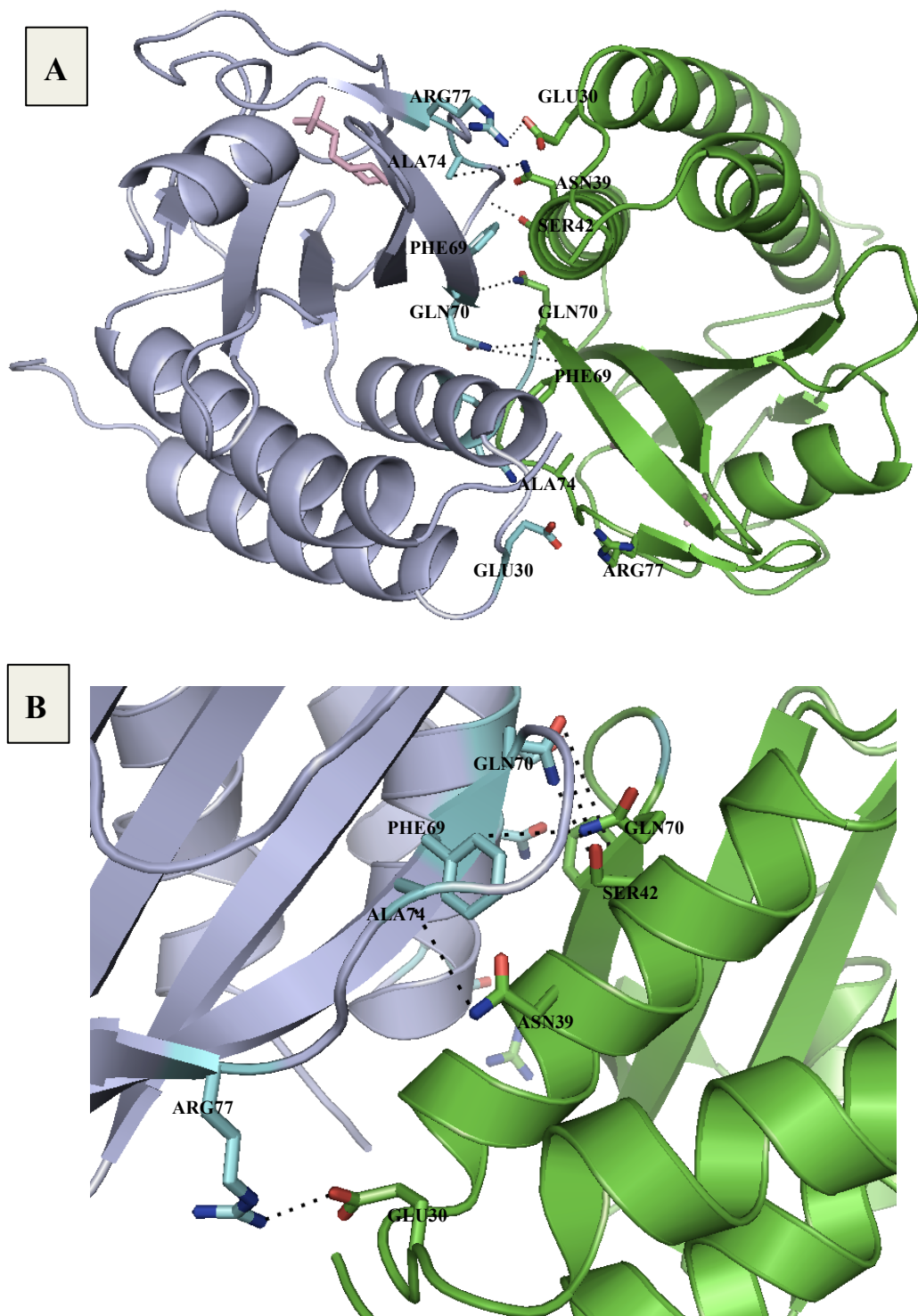


Figure 6.22: Residues positioned in BPSL2418_{MES} monomer-monomer interface. The residues highlighted by the PISA analysis are shown in stick form. (A) Hydrogen bonds and salt bridges involved in the interface shown as dashed lines. (B) Close view of the dimer interface. The two MES molecules of the two monomers are shown in light pink color. The images were created by PyMol (DeLano & Lam, 2005).

6.2.2 BPSL2418_{Met-SO} analysis

The PISA webservice was used with a model containing two molecules of BPSL2418_{Met-SO} each binding a substrate Met-SO molecule in the asymmetric unit.

6.2.2.1 BPSL2418_{Met-SO} monomers

The two molecules of BPSL2418_{Met-SO} were labeled as chain A and chain B. Chain A contains 152 residues, 132 of them are surface exposed residues as defined by PISA. It has a solvent accessible area of approximately 7100 Å² and estimated solvation free energy gain upon formation of the interface (ΔG) of -144.8 Kcal/mol. Chain B has 133 residues and there are surface exposed residues. The solvent accessible area for chain B is approximately 7000 Å² and the solvation free energy gain (ΔG) is estimated as -146.4 Kcal/mol.

6.2.2.2 Monomer-monomer interface forming dimeric BPSL2418_{Met-SO}

PISA analysis for BPSL2418_{Met-SO} predicts an interaction between the two molecules (A and B) in the asymmetric unit forming a dimeric state with CSS = 1. It shows that the crystal packing appears between BPSL2418_{Met-SO} molecules related by a 2-fold axis (Figure 6.23). PISA analysis does not suggest any interaction between the asymmetric unit dimers to form tetramers. The dimer interface has a total of 25 residues from each molecule (table 6.6) and a buried surface area of 850 Å² and it contains 15 hydrogen bonds and 8 salt bridges (table 6.7). The interface area approximately includes 12% of the total surface area of BPSL2418_{Met-SO} molecule, and this percentage confirms the dimeric state of the protein (Jones & Thornton, 1995). Also the gel filtration analysis (Figure 4.10) agrees with the PISA prediction. BPSL2418_{Met-SO} dimeric interface (Figure 6.24) is very similar to the interface of BPSL2418_{MES} dimer except a HIS46[NE2]-GLY71[O] hydrogen bond occurs in BPSL2418_{Met-SO} dimer, where BPSL2418_{MES} dimer has PHE69[O]-GLN70[NE2] hydrogen bond instead. PHE69 and GLN70 have no changes in BPSL2418_{Met-SO} which may be suggested the formation of a

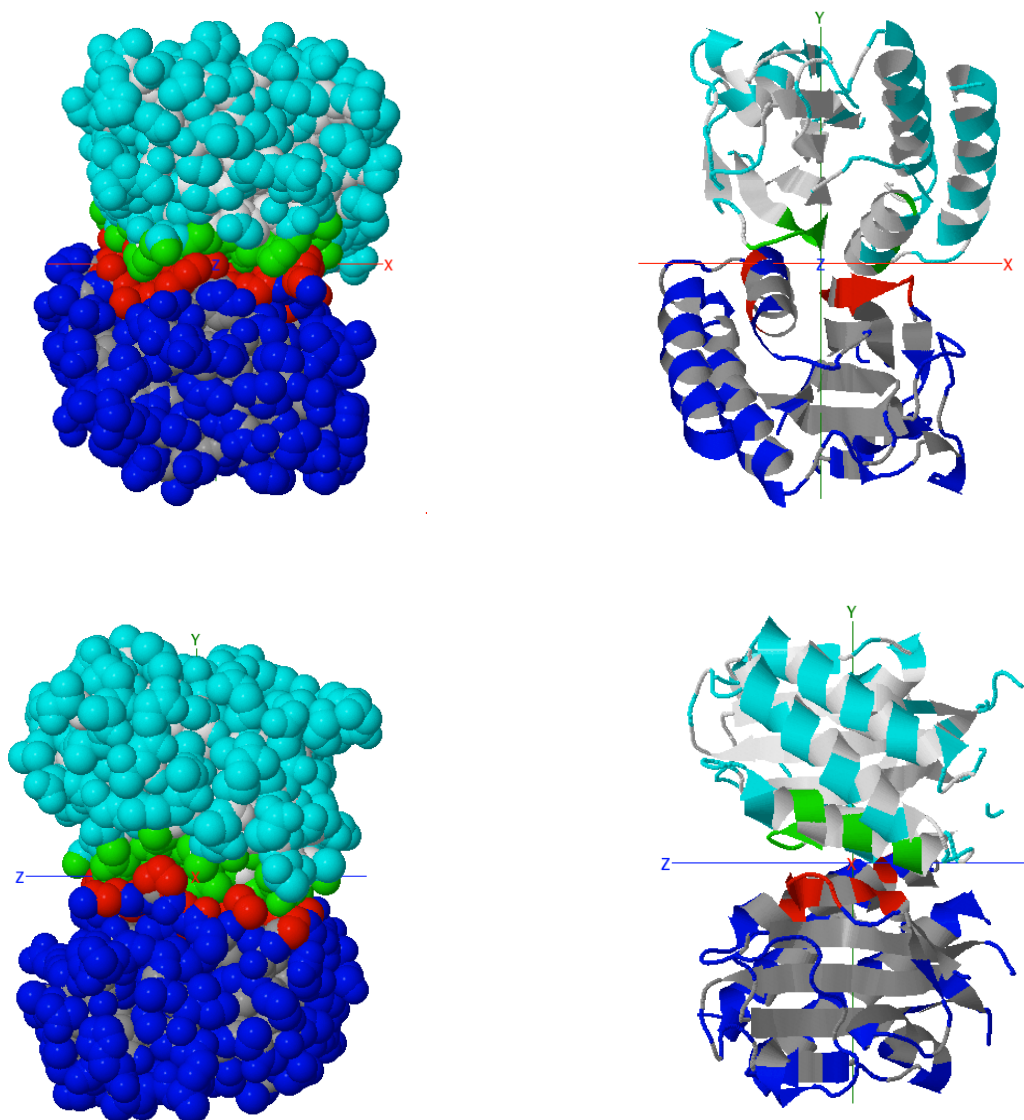


Figure 6.23: The dimeric structure model of BPSL2418_{Met-SO}. On the left side the space-filling model and on the right side the cartoon model, each has three images taken 90° apart about the X-axis. The two molecules forming the dimer are in blue and tiffany colors and the interface area in green and red colors. The Figure was produced using PISA webserver <http://pdbe.org/PISA>.

	Residue	Accessible		Buried			
		surface area (Å ²)	area Monomer I	surface area (Å ²)	area Monomer I	surface area (Å ²)	area Monomer II
1	LEU 26	63.88		19.43		70.64	15.96
2	GLU 30	65.12		49.60		76.10	63.97
3	THR 31	87.61		5.64		81.79	7.24
4	ASP 32	74.33		21.41		72.20	20.94
5	VAL 34	91.96		56.24		86.28	56.56
6	ALA 35	41.67		41.67		40.91	40.91
7	ASN 36	3.06		1.60		1.72	1.35
8	ALA 38	43.27		41.44		44.60	41.56
9	ASN 39	77.40		75.87		75.00	73.96
10	SER 42	37.00		37.00		34.15	34.15
11	LEU 43	38.80		14.97		39.88	13.82
12	HIS 46	130.72		46.89		131.40	49.62
13	PHE 58	42.30		0.47		39.34	1.72
14	VAL 65	36.79		28.95		33.37	25.59
15	VAL 66	32.68		30.44		34.64	32.19
16	GLY 67	17.52		16.39		16.94	16.13
17	PRO 68	29.03		27.69		26.45	25.77
18	PHE 69	59.61		55.35		63.76	59.15
19	GLN 70	54.28		48.15		44.26	40.18
20	GLY 71	36.06		31.02		41.49	36.67
21	LYS 72	148.40		28.78		150.48	30.90
22	PRO 73	123.35		77.27		117.53	74.766
23	ALA 74	23.22		10.23		22.81	8.30
24	VAL 76	90.21		51.55		84.79	54.58
25	ARG 77	127.05		38.36		132.00	36.16

Table 6.6: Accessibility and buried surface areas of 25 residues involved in the BPSL2418_{Met-SO} monomer-monomer interface calculated by PISA <http://pdbe.org/PISA> (Krissinel & Henrick, 2007).

Hydrogen bonds

	Monomer I	Distance (Å)	Monomer II
1	ARG 77 [NE]	2.94	GLU 30 [OE1]
2	ARG 77 [NH2]	2.85	GLU 30 [OE2]
3	ALA 74 [N]	2.87	ASN 39 [OD1]
4	GLY 71 [N]	3.81	GLN 70 [OE1]
5	HIS 46 [NE2]	3.16	GLY 71 [O]
6	SER 42 [OG]	2.68	GLY 71 [O]
7	GLN 70 [NE2]	3.36	GLY 71 [O]
8	ASN 39 [ND2]	3.34	ALA 74 [O]
9	GLU 30 [OE1]	2.87	ARG 77 [NE]
10	GLU 30 [OE2]	2.82	ARG 77 [NH2]
12	ASN 39 [OD1]	3.12	ALA 74 [N]
13	GLY 71 [O]	2.65	SER 42 [OG]
14	GLY 71 [O]	3.44	GLN 70 [NE2]
15	ALA 74 [O]	3.02	ASN 39 [ND2]

Salt bridges

	Monomer I	Distance (Å)	Monomer II
1	ARG 77 [NE]	3.93	GLU 30 [OE1]
2	ARG 77 [NH]	2.94	GLU 30 [OE1]
3	ARG 77 [NH2]	2.85	GLU 30 [OE2]
4	ARG 77 [NE]	3.40	GLU 30 [OE2]
5	GLU 30 [OE1]	2.87	ARG 77 [NE]
6	GLU 30 [OE1]	3.71	ARG 77 [NH2]
7	GLU 30 [OE2]	3.53	ARG 77 [NE]
8	GLU 30 [OE2]	2.82	ARG 77 [NH2]

Table 6.7: The hydrogen bonds and salt bridges formed in BPSL2418_{Met-SO} monomer-monomer interface calculated by PISA <http://pdbe.org/PISA> (Krissinel & Henrick, 2007).

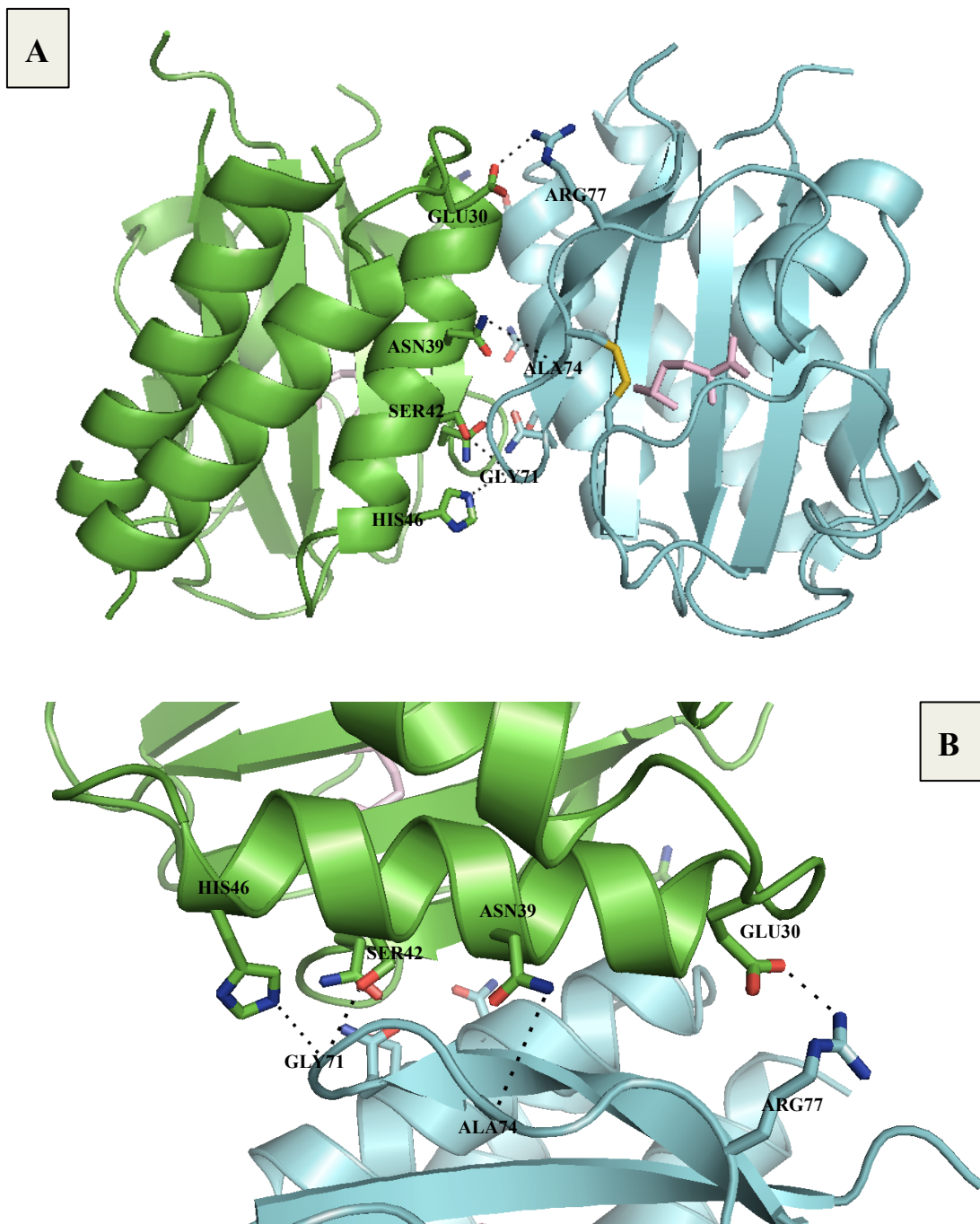


Figure 6.24: Residues positioned in BPSL2418_{Met-SO} monomer-monomer interface.

The residues highlighted by the PISA analysis are shown in stick form (A) Hydrogen bonds and salt bridges involved in the interface shown as dashed lines. (B) Close view of the dimer interface. The two Met-SO molecules of the two monomers are shown in pink color and the CYS75-CYS109 disulfide is shown in yellow. The images were created by PyMol (DeLano & Lam, 2005).

hydrogen bond between these two residues occasionally. Similar to BPSL2418_{MES}, the dimer interface of BPSL2418_{Met-SO} consists of residues in $\alpha 2$, $\alpha 1$ - $\alpha 2$ loop, $\beta 2$ and $\beta 2$ - $\beta 3$ loop interact with residues in $\alpha 2$, $\alpha 1$ - $\alpha 2$ loop, $\beta 2$ and $\beta 2$ - $\beta 3$ loop of the symmetry related molecule.

6.2.3 BPSL2418_{Reduced} analysis

A model of BPSL2418_{Reduced} structure containing 2 molecules in the asymmetric unit was used to run PISA analysis.

6.2.3.1 BPSL2418_{Reduced} monomers

Chain A and chain B of BPSL2418_{Reduced} dimer have 154 and 152 residues respectively. There are 136 and 133 surface exposed residues are 136 residues from chain A and 133 from chain B, with solvent accessible area of 7300 Å² and 7100 Å² and estimated solvation free energy gain upon formation of the interface (ΔG) -140.6 Kcal/mol and -143.4 Kcal/mol respectively.

6.2.3.2 Monomer-monomer interface forming dimeric BPSL2418_{Reduced}

PISA analysis indicates a dimeric state with CSS equal to 0.75, which also agreed with gel filtration (Figure 4.10). The dimeric interface includes 25 residues from chain A and 24 residues from chain B (table 6.8), a buried surface area of 830 Å², and it contains 15 hydrogen bonds and 8 salt bridges (table 6.9). The interface area presents 12% of the total surface area of the protein molecule. This interface has PHE69[O]-GLN70[NE2] similar to BPSL2418_{MES} dimer. Similar to BPSL2418_{MES} and BPSL2418_{Met-SO} the dimer interface of BPSL2418_{Reduced} consists of residues in $\alpha 2$, $\alpha 1$ - $\alpha 2$ loop, $\beta 2$ and $\beta 2$ - $\beta 3$ loop interact with residues in $\alpha 2$, $\alpha 1$ - $\alpha 2$ loop, $\beta 2$ and $\beta 2$ - $\beta 3$ loop of the symmetry related molecule. The images of BPSL2418_{Reduced} dimer produced using PISA are identical to the images of BPSL2418_{Met-SO} dimer (Figure 6.25).

	Residue	Accessible		Buried	
		surface area (Å ²)	Monomer I	surface area (Å ²)	Monomer I
1	LEU 26	73	19.00	75.24	14.34
2	GLU 30	66.62	52.32	73.91	57.59
3	THR 31	83.95	5.30	80.69	10.55
4	ASP 32	71.66	21.59	71.82	19.78
5	VAL 34	90.96	59.26	88.43	56.05
6	ALA 35	39.90	39.58	39.72	39.72
7	ASN 36	3.22	0.74	3.64	1.56
8	ALA 38	37.99	37.15	42.34	39.42
9	ASN 39	82.92	70.72	81.93	77.84
10	SER 42	34.50	34.50	31.49	31.49
11	LEU 43	38.70	13.10	37.31	14.63
12	HIS 46	131.48	26.77	138.53	56.16
13	PHE 58	43.39	0.63	40.95	2.34
14	VAL 65	35.13	25.60	35.15	25.73
15	VAL 66	27.62	26.11	31.68	30.57
16	GLY 67	21.27	19.46	16.80	15.98
17	PRO 68	29.03	27.70	27.25	26.42
18	PHE 69	63.12	58.48	60.53	57.69
19	GLN 70	46.89	43.56	51.67	45.59
20	GLY 71	45.45	40.90	54.12	40.05
21	LYS 72	163.59	32.77	-	-
22	PRO 73	127.49	71.70	137.04	67.76
23	ALA 74	27.84	8.50	25.41	8.95
24	VAL 76	78.89	52.59	87.71	53.89
25	ARG 77	133.31	36.84	129.81	35.99

Table 6.8: Accessibility and buried surface areas residues involved in the BPSL2418_{Reduced} monomer-monomer interface calculated by PISA <http://pdbe.org/PISA> (Krissinel & Henrick, 2007).

Hydrogen bonds

	Monomer I	Distance (Å)	Monomer II
1	ARG 77 [NE]	2.96	GLU 30 [OE1]
2	ARG 77 [NH2]	3.12	GLU 30 [OE2]
3	ARG 77 [NH2]	3.74	THR 31 [OG1]
4	ALA 74 [N]	2.94	ASN 39 [OD1]
5	GLN 70 [NE2]	2.90	PHE 69 [O]
6	GLY 71 [N]	3.29	GLN 70 [OE1]
7	SER 42 [OG]	2.69	GLY 71 [O]
8	ASN 39 [ND2]	3.19	ALA 74 [O]
9	GLU 30 [OE1]	2.83	ARG 77 [NE]
10	GLU 30 [OE2]	2.70	ARG 77 [NH2]
11	ASN 39 [OD1]	3.15	ALA 74 [N]
12	PHE 69 [O]	3.03	GLN 70 [NE2]
13	GLN 70 [OE1]	3.22	GLY 71 [N]
14	GLY 71 [O]	2.54	SER 42 [OG]
15	ALA 74 [O]	3.02	ASN 39 [ND2]

Salt bridges

	Monomer I	Distance (Å)	Monomer II
1	ARG 77 [NE]	3.94	GLU 30 [OE1]
2	ARG 77 [NE]	2.96	GLU 30 [OE1]
3	ARG 77 [NH2]	3.12	GLU 30 [OE2]
4	ARG 77 [NE]	3.59	GLU 30 [OE2]
5	GLU 30 [OE1]	2.83	ARG 77 [NE]
6	GLU 30 [OE1]	3.75	ARG 77 [NH2]
7	GLU 30 [OE2]	3.36	ARG 77 [NE]
8	GLU 30 [OE2]	2.70	ARG 77 [NH2]

Table 6.9: The hydrogen bonds and salt bridges formed in BPSL2418_{Reduced} monomer-monomer interface calculated by PISA <http://pdbe.org/PISA> (Krissinel & Henrick, 2007).

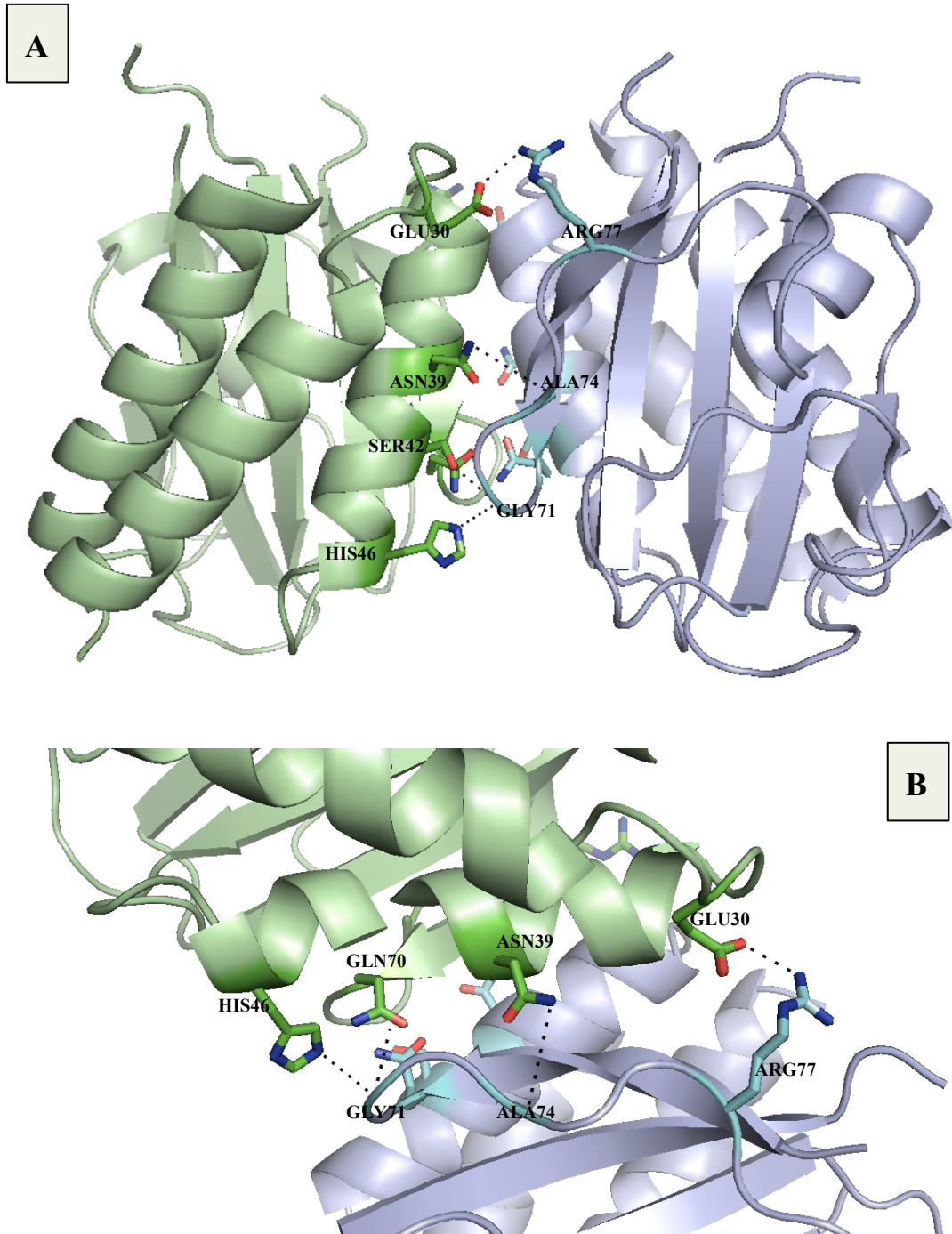


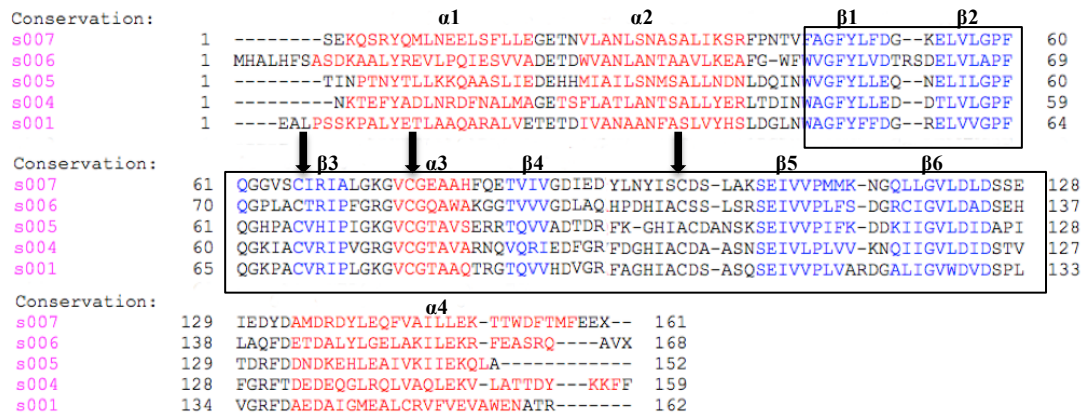
Figure 6.25: Residues positioned in BPSL2418_{Reduced} monomer-monomer interface. The residues highlighted by the PISA analysis are shown in stick form (A) Hydrogen bonds and salt bridges involved in the interface shown as dashed lines. (B) Close view of the dimer interface. The images were created by PyMol (DeLano & Lam, 2005).

6.3 Functional prediction

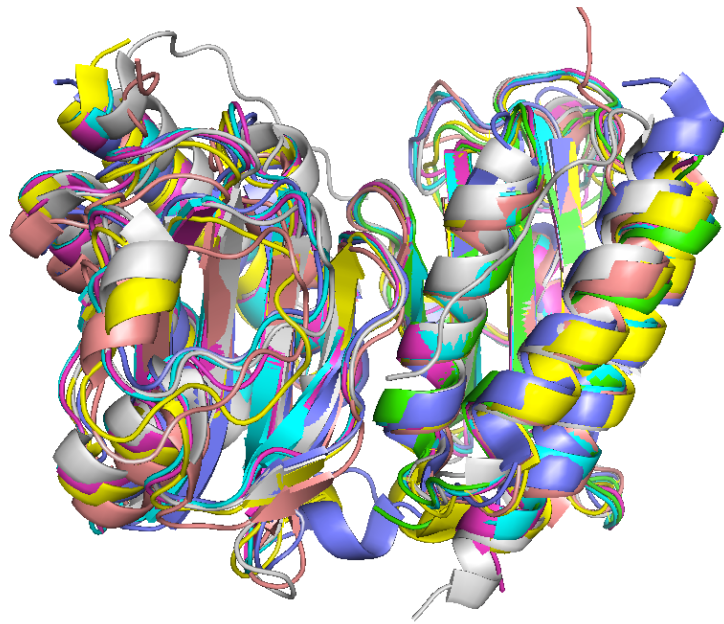
The structures of the BPSL2418 forms can be used to predict the function of the BPSL2418 protein. The determination of the protein family that BPSL2418 belongs to and the high identity of BPSL2418 with known function homologs can provide useful information for understanding the protein.

6.3.1 Structure comparisons

All the final refined structures of BPSL2418 forms were analyzed with the Dali server as described in section 6.1. The Dali search shows that BPSL2418 belongs to the GAF domain free methionine sulfoxide reductase (fRMsR) family. In the database only four structures of fRMsR are known so far: *Escherichia coli* fRMsR (PDB code 1VHM), *Staphylococcus aureus* fRMsR (PDB code 3KSF), *Neisseria meningitidis* fRMsR (PDB code 3MMH) and *Streptococcus pneumoniae* fRMsR (PDB code 3RFB). As demonstrated in section 6.1, each structure of the BPSL2418 forms superimposed very well with structures from the fRMsR family. However in order to obtain the sequence similarity and conserved residues between BPSL2418 and fRMsR proteins, they were also aligned using PROMALS3D (Pei *et al.*, 2008). The alignment was run using the 3D models of these proteins (Figure 6.26). 12 residues of 13 residues composing the active site are conserved among the aligned proteins. The three cysteines in loop1, $\alpha 3$ and loop4 are conserved where a disulfide bond is formed between the cysteine in loop1 and the cysteine in loop4. The third cysteine located in $\alpha 3$ provides a hydrogen bond via its main chain amide nitrogen in the ligand interaction (see section 6.2). Like all GAF domains, BPSL2418 and the fRMsR known structures are dimeric that also superimposed very well (Figure 6.27).



6.26: Structure based alignment of BPSL2418 (s001), *N. meningitidis* fRMs (s004), *S. pneumoniae* fRMs (s005), *E. coli* fRMs (s006) and *S. aureus* fRMs (s007). The β -strands and α -helices are represented as blue and red residues, respectively. The active site consisting of (β 1, β 2, loop1, β 3, loop2, α 3, loop3, β 4 and loop4) is defined by black rectangles and the three conserved cysteines are indicated by black arrows.



6.27: structural alignment of BPSL2418 structure and fRMsR structures (*N. meningitidis* fRMsR *S. pneumoniae* fRMsR, *E. coli* fRMsR and *S. aureus* fRMsR). The Figure shows an alignment of dimeric structures of BPSL2418 with fRMsR proteins. The images were created by PyMol (DeLano & Lam, 2005).

6.3.2 Catalytic Mechanism of BPSL2418

All the structural analyses of BPSL2418 indicate that this protein is most likely a new member in the GAF domain fRMs_r family. The crystal structures of BPSL2418_{MES}, BPSL2418_{Met-SO} and BPSL2418_{Reduced} help in describing the role of BPSL2418 in reversing the oxidation of the free methionine residue, in understanding the binding method of Met-SO substrate to the active site, and the conformational changes that occur during the reduction. The crystal structures of the BPSL2418 forms indicate significant conformational changes in the active site in each catalytic step. The changes in the active site occur particularly in loop4 including the catalytic CYS109. The reduced form (BPSL2418_{Reduced}) creates an opened conformation (Figure 6.28) mostly to permit admission to the substrate. In the Met-SO substrate binding form, BPSL2418_{Met-SO} has a closed conformation after access of the substrate (Figure 6.28). The comparison between the BPSL2418 forms revealed that the cysteine located in loop4 CYS109 was the most mobile while the CYS75 position in loop1 and CYS85 in α 3 remain unchanged (Figure 6.17), but CYS85 may play a role in substrate binding. This is in agreement with the *Staphylococcus aureus* fRMs_r (Bong *et al*, 2010) and the *Neisseria meningitidis* fRMs_r (Gruez *et al*, 2010) where the structural and biochemical analyses of these proteins suggest that their catalytic cysteines are CYS102 and CYS118, respectively which correspond to CYS109 in BPSL2418. CYS102 and CYS118 are the mobile cysteines, which form a disulfide bond with CYS68 and CYS84, respectively leading to conformational changes in the active site (see section 1.2.3 (2) and (3)). The corresponding residues to CYS75 in *Staphylococcus aureus* fRMs_r and the *Neisseria meningitidis* fRMs_r are CYS68 and CYS84 respectively and they are immobile cysteines as CYS75 in BPSL2418. The third conserved cysteine CYS78 and CYS94 located in α 3 of *Staphylococcus aureus* fRMs_r and the *Neisseria meningitidis* fRMs_r respectively are also immobile cysteines as BPSL2418. The crystal structure of BPSL2418 in complex with Met-R-SO illuminates not only the catalysis of the reductase step but also how BPSL2418 binds Met-R-SO with a strong preference for the Met-R-SO isomer. The structure of the reduced form has an open conformation to allow access to the substrate. Another complex structure BPSL2418_{MES} was determined in this project with a MES molecule in the binding site. BPSL2418_{MES} is an oxidized form with a disulfide bond between CYS75 and

CYS109. This suggests that MES molecule can act as competitive inhibitor of BPSL2418 enzyme. Our structural analysis suggests that CYS109 is the catalytic residue of BPSL2418. The hydrophobic region in the active site could accommodate the ϵ -methyl group of the substrate via van der Waals interactions, where the hydrophilic region with the nitrogen of the substrate. Also, this hydrophilic region may play a role in stabilizing the protonated oxygen atom of the sulfoxide moiety during catalysis. Thus, the hydrophobic affinity to substrate, whereas the hydrophilic region seems important for binding specificity.

Taken together, it can be suggested that the catalytic mechanism of BPSL2418 includes three steps (Figure 6.29):

- (1) The BPSL2418 is reduced by a reductant (usually Trx in vivo or DTT in vitro) generating an opened pocket to permit access to the free Met-SO substrate. CYS109 in BPSL2418 loop4 attacks the sulfoxide group of Met-SO and is then oxidized forming CYS sulfenic acid (CYS-SOH). This reverses the Met-SO into a free methionine residue.
- (2) CYS75 in BPSL2418 loop1 interacts with the CYS109 sulfenic acid intermediate forming a disulfide bond (generating a close conformation) and releasing a water molecule.
- (3) The BPSL2418 enzyme activity is regenerated by reducing the CYS75-CYS109 disulfide.

The crystals of the different forms of BPSL2418 comprise dimer in the asymmetric unit and the two active sites within the dimer are separated from each other by approximately 23.8 Å, which suggests no communication between the two active sites. Also the interfaces have no significant change seen between the structures. In thermophilic enzymes, a decrease in flexibility cause low enzyme activity while in less stable proteins an increase in flexibility is associated with enhanced enzyme activity. In this study the role of active site rigidity in enzyme activity is misleading because the disulfide bond forms after Met-SO reduction. Thioredoxin reductase is the only enzyme known to catalyze the reduction of thioredoxin and hence is a central component in the thioredoxin system. Together with thioredoxin (Trx) and NADPH this system's most general description is as a method of forming reduced disulfide bonds in cells. Electrons are taken from NADPH via TrxR and are

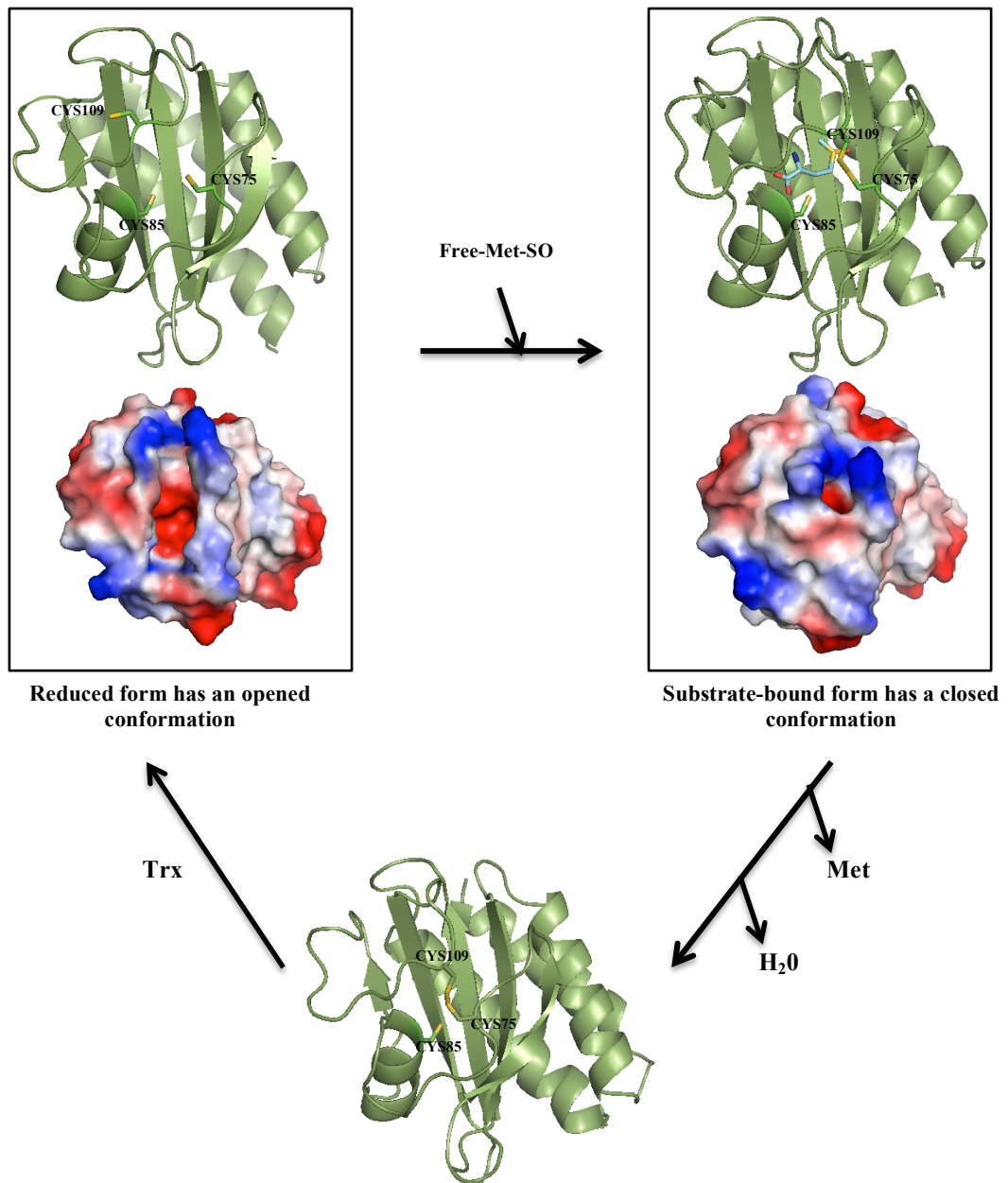


Figure 6.28: Structural representation of proposed catalytic mechanism of BPSL2418. Reduced BPSL2418 exhibits an opened conformation permitting the access of Met-SO into the active site. After substrate binding, the BPSL2418 enzyme forms a disulfide bond and is converted to a closed conformation.

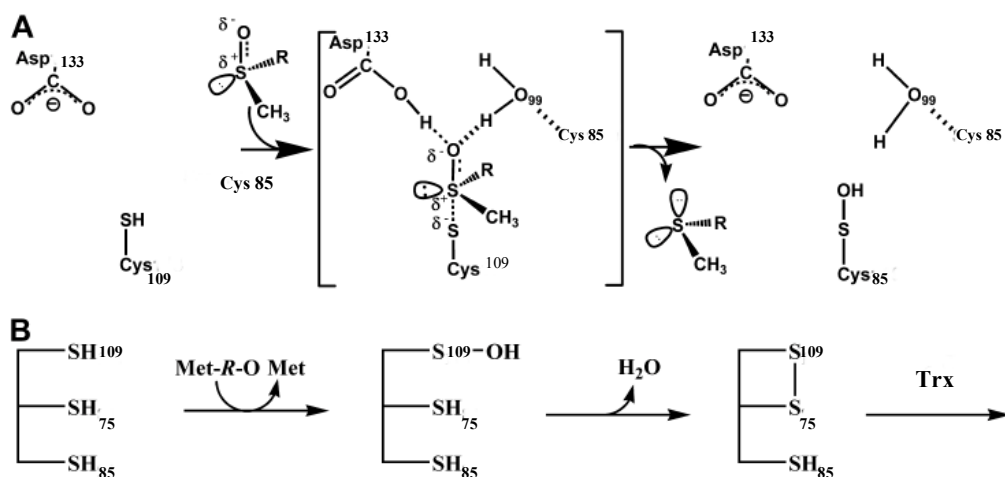


Figure 6.29: Schematic representation of proposed catalytic mechanism of BPSL2418. (A) The substrate binds to the active site with its sulfoxide function largely polarized, leading to a transition state compatible with a sulfurane of bipyramidal geometry. The rearrangement of the transition state leads to the formation of a sulfenic acid intermediate on the catalytic Cys₁₀₉ residue. In this scenario, Asp₁₃₃ and Cys₈₅, stabilize the sulfurane transition state via interaction with the OH group. (B) The reductase step (the first one) leads to the formation of a sulfenic acid intermediate on the catalytic Cys₁₀₉ and of Met. In the second step, attack of Cys₇₅ on the sulfur atom of the sulfenic acid intermediate leads to the formation of a disulfide bond between Cys₁₀₉ and Cys₇₅ and the release of a water molecule. Return of the active site to a fully reduced state can proceed via reduction of the Cys₁₀₉–Cys₇₅ disulfide bond formed by Trx.

transferred to the active site of Trx, which goes on to reduce protein disulfides or other substrates. The Trx system exists in all living cells and has an evolutionary history tied to DNA as a genetic material, defense against oxidative damage due to oxygen metabolism, and redox signaling using molecules like hydrogen peroxide and nitric oxide (Meyer et al, 2009). However, other GAF domain proteins basically play role in ligand binding specifically cyclic nucleotide for signal potentiation. This may suggest that Met-SO plays role as a signaling ligand in response to oxidative stress in some organism. This might be in a signaling cascade which is a series of chemical reactions that are initiated by a first messenger acting on a receptor that is transduced to the cell interior through second messengers (which amplify the initial signal) and ultimately to effector molecules, resulting in a cell response to the initial stimulus.

6.4 Conclusion and future work

The methionine sulfoxide reductase (Msr) is the responsible protein family of the reduction of methionine sulfoxide (Met-SO). The Msr family includes three types: MsrA, MsrB and fRMsr. MsrA and MsrB capable of reducing free Met-SO and Met-SO in peptides. This mechanism is stereospecific with the *S*- and *R* forms of Met-SO reduced by MsrA and MsrB, respectively. The third type of Msr is the free methionine-*R*-sulfoxide reductase (fRMsr), which catalyzes the reduction of the *R*-form of free methionine sulfoxide back to free methionine. The fRMsr subfamily discovered in 2007 and interestingly only four structures of this type are known so far. The fRMsr type belongs to the GAF domain family and it is the only member that shows enzymatic activity. Other GAF domain proteins basically play role in ligand binding specifically cyclic nucleotide for signal potentiation. This may suggest that Met-SO plays role as a signaling ligand in response to oxidative stress in some organism. This thesis presents the BPSL2418 protein from *Burkholderia pseudomallei* as a novel member of fRMsr family. The crystal structures of the reduced (BPSL2418_{Reduced}), Met-SO-bound (BPSL2418_{Met-SO}) and MES-bound (BPSL2418_{MES}) forms of free methionine-*R*-sulfoxide reductase from *Burkholderia pseudomallei* were determined in this work. This is the first project provides the structural analysis of MES-bound form with the reduced and substrate-bound forms

of a free methionine-*R*-sulfoxide reductase protein. The structural analysis and comparison of the BPSL2418 forms indicate conformational changes in the active site particularly in loop4 (residues 98-113). The project suggests the catalytic mechanism of BPSL2418 in which CYS109 (located in loop4) is a mobile residue functions as the catalytic cysteine and CYS75 functions as the resolving cysteine, which forms a disulfide bond with CYS109. The position of CYS85 is like CYS75 remains unchanged during the catalytic activity; CYS85 involves in the interaction of BPSL2418 with Met-SO substrate and MES molecule.

In this project future work should focus on produce the crystal structure of the unbound oxidized form to complete the comparison image of the reduction mechanism. Also obtain a crystal structure of CYS109, CYS102 and CYS85 mutant forms either as a single or double mutations can confirm the catalytic residues and the function of each cysteine. In this project no experiment have done to test the substrate specificity of BPSL2418 enzyme which suggest performing an enzyme assay to monitor the reduction activity against different substrates such as free and peptide bound forms of *S*- isomer Met-SO and the peptide bound form of *R*- isomer Met-SO. Also enzyme assays can be used to examine the activity of BPSL2418 mutants comparing with the wild type. This can be achieved by using steady-state kinetics of wild type and mutants with Trx-regeneration system and with Met-*R*-SO as a substrate and Trx as a reductant.

Chapter 7: CT1 and CT2 Expression, Purification and crystallization attempts

This chapter presents all the efforts to express, purify and crystallize the C-terminal domains of Polycystin1 (CT1) (26 kDa) and of Polycystin2 (CT2) (35 kDa) in order to use the crystals of these proteins for experimental X-ray diffraction.

7.1 Preparation of CT1 protein

7.1.1 Recombinant plasmid of CT1

The insert gene including a GB1 domain at the N-terminus was ligated into the *E. coli* expression vector pET-21a(+) which contains an ampicillin resistance gene. The GB1 domain is an immunoglobulin-binding fusion protein (8 kDa) often used to increase the solubility of other proteins (Cheng & Patel, 2004). The GB1-CT1 pET-21a(+) plasmid was provided by Prof. Albert Ong.

7.1.2 CT1 overexpression in *E. coli*

An *E. coli* expression method was used to express GB1-CT1. To achieve large-scale 4 l expression the following protocol was applied: competent cells (BL21-DE3-RIPL) were transformed with the GB1-CT1pET-21a(+) construct. Cells were plated on LB-agar plates provided with 75 µg/ml Chloramphenicol to select BL21-DE3-RIPL, 50 µg/ml ampicillin to select the plasmid and 1% glucose to inhibit expression. LB-agar plates were incubated at 37 °C overnight. A starter culture of LB (5 ml) supplemented with 50 µg/ml ampicillin and 1% glucose was inoculated with a single colony and incubated at 37 °C, 200 rpm overnight. The secondary culture of LB (500 ml) supplemented with 50 µg/ml ampicillin, 75 µg/ml Chloramphenicol and 1% glucose was inoculated at 1:100 dilution of the starter culture and incubated at 37 °C, 200 rpm until the OD_{600 nm} reached 0.6-0.9

(approximately 4-5 hours). CT1 expression was induced by adding 1 mM IPTG to the culture. The main cultures were incubated at 37 °C, 200 rpm for 3 hours. Cells were harvested by centrifugation at 6000 g for 10 minutes. Pellets were stored at -80 °C. GB1-CT1 was expressed as insoluble inclusion bodies (Figure 7.1). Expression conditions were optimized to try to achieve soluble yield, but no soluble protein was expressed. The optimization included changing induction time, induction temperature and IPTG concentration. None of the optimization attempts produced soluble protein.

7.1.3 CT1 protein refolding

The mentioned protocol in section 2.12 was used to refold the inclusion bodies of CT1, this attempt did not improve the protein folding.

7.2 Preparation of CT2 protein

7.2.1 Recombinant plasmid of CT2

The human polycystin-2 C-terminal domain (CT2) was amplified by the PCR technique in a His-CT2 construct. The insert gene included a His tag at the N-terminus. The PCR product was ligated into the *E. coli* expression vector pET-28a(+) which contains a kanamycin resistance gene (Figure 7.2). This recombinant plasmid was provided by Prof. Albert Ong.

7.2.2 CT2 overexpression in *E. coli*

An *E. coli* expression method was used to express His-CT2 protein. To achieve large-scale 4 l expression the following protocol was applied: competent cells (BL21-DE3-RIPL) were transformed with the His-CT2 pET-28a(+) construct. Cells were plated on LB-agar plates provided with 75 µg/ml chloramphenicol to select BL21-DE3-RIPL, 50 µg/ml kanamycin to select the plasmid and 1% glucose. LB-agar plates were incubated at 37 °C overnight. A starter culture of LB (5 ml)

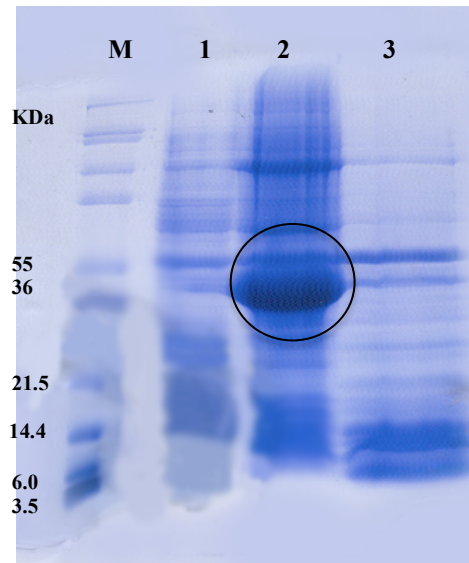
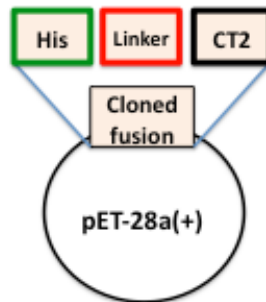


Figure 7.1: 12% SDS- PAGE gel showing GB1-CT1 (34 kDa) expression. Lane M: Molecular weight marker, lane 1: Pre-induction, lane 2: Band in the insoluble fraction and lane 3: No bands in the soluble fraction.



```

mgsshhhhh ssglvprgsh Mindtysevk sdlaqqkaem
elsdlirkgy hkalvklklk kntvddises lrqgggklnf
delrqdlkgk ghtdaeieai ftkydqgdgq eltehehqqm
rddlekered ldldhsslpr pmssrsfprs lddseeddde
dsghssrrrg sissgvsyee fqvlvrrvdr mehsigsivs
kidavivkle imeraklkr evlgrlldgv aederlgrds
eihreqmerl vreelerwes ddaasqishg lgtpvglngq
prprssrpss sqstegmega ggngssnvhv

```

Figure 7.2: His-CT2 construct (35 kDa). Green color indicates the histidine tag, red color indicates linker amino acids including thrombin site in blue (rgs). CT2 residues colored in black. Another thrombin site in the middle of CT2 sequence(rgs).

supplemented with 50 $\mu\text{g/ml}$ kanamycin and 1% glucose was inoculated with a single colony and incubated at 37 °C, 200 rpm for overnight. The secondary culture of LB (500 ml) supplemented with 50 $\mu\text{g/ml}$ kanamycin, 75 $\mu\text{g/ml}$ chloramphenicol and 1% glucose was inoculated at 1:100 dilution of the starter culture and incubated at 37 °C, 200 rpm until the OD_{600 nm} reached 0.6-0.9. (approximately 2-5 hours). CT2 expression was induced by adding 1 mM IPTG to the culture. The main cultures were incubated at 37 °C, 200 rpm for 3 hours. Cells were harvested by centrifugation at 6000 g for 10 minutes. Pellets were stored at -80 °C. CT2 was expressed by the methods described above and the SDS-PAGE gel reveals a strong expression band in the soluble fraction in the expected size (Figure 7.3).

7.2.3 CT2 purification

The cell pellet from the large-scale *E. coli* expression was defrosted on ice, and was resuspended in lysis buffer (50 mM Tris pH8, 500 mM NaCl) using a 1:5 w/v ratio. The suspension was incubated on ice for 30 minutes then it was aliquoted into 2 centrifuge tubes (10 ml in each tube). The suspension in each tube was sonicated for 3x20 seconds (on ice) and centrifuged at 40,000 g at 4 °C for 30 minutes. The supernatant was taken for the purification step. Soluble fractions of His-CT2 were purified by Nickel column affinity chromatography. The purity of eluted His-CT2 protein was checked by 12% SDS-PAGE. The His-CT2 purification using the Nickel column was run successfully and bands were seen on the gel at the expected size (Figure 7.4). The SDS-PAGE showed many impurities, which suggested another purification step was needed. Gel filtration chromatography was used as a second purification step. Gel filtration was performed on a Hi load Superdex200, 16x60 cm (GE healthcare life science) column. The column flow rate was 1.5 ml/min. Gel filtration was performed in gel filtration buffer (25 mM Tris, 150 mM NaCl, pH 8.0 at 4 °C) and the purity of eluted protein was checked by 12% SDS-PAGE (Figure 7.5 a). The gel filtration elution profile shows that CT2 tends to form oligomers (Figure 7.5 b). The calibration plot for the Hi-load Superdex 200 column was used to calculate the protein molecular weight and this gave a value of approximately 350 kDa, which

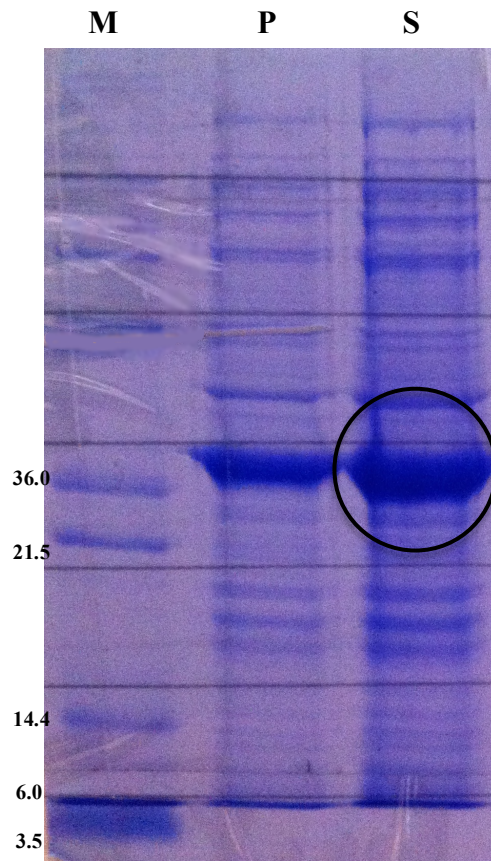


Figure 7.3: 12% SDS- PAGE gel showing His-CT2 (35 KDa) overexpression. Lane M: Molecular weight marker. Lane P: Insoluble fraction. Lane S: Soluble fraction

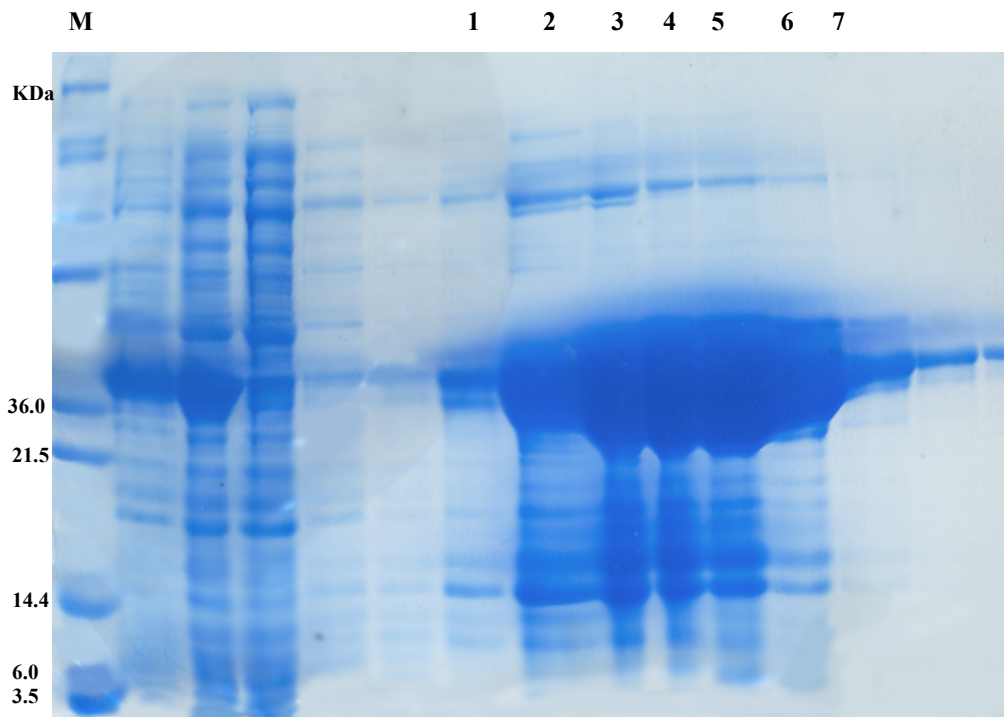


Figure 7.4: 12% SDS-PAGE of His-CT2 (35 KDa) nickel affinity chromatography. Lane M: Molecular weight marker . Lane 1-7 nickel column fractions.

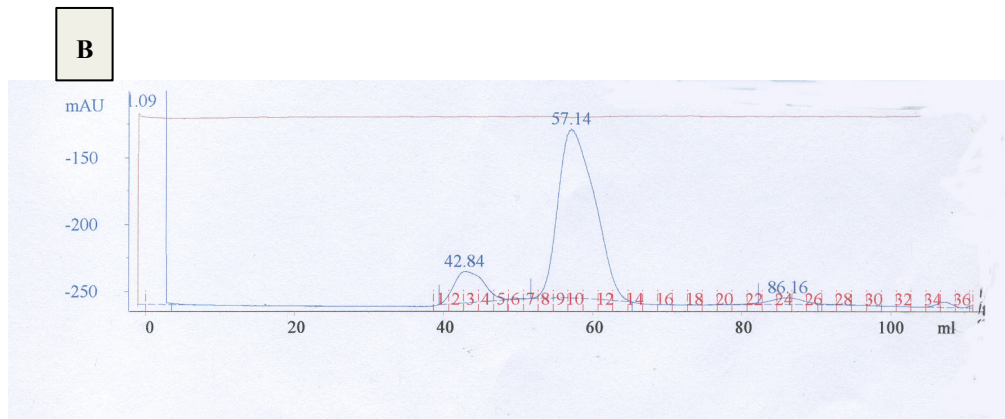
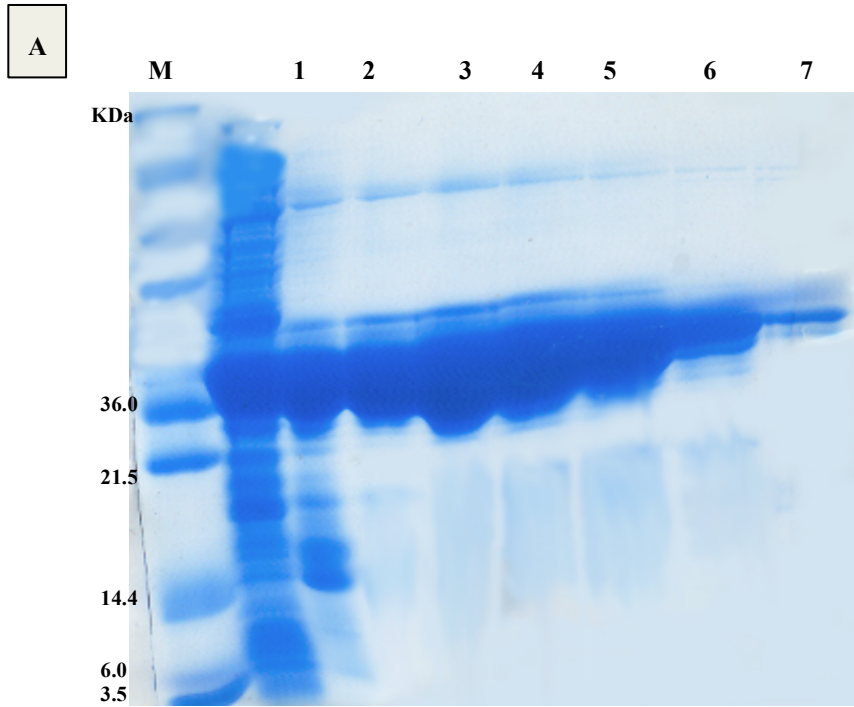


Figure 7.5: Gel filtration of His-CT2 (35 KDa). A: 12% SDS-PAGE of samples 2-8 gel filtration chromatography. B: Gel filtration elution profile (peak 57.14 ml).

would correspond roughly to an octamers (Figure 7.6). The protein concentration was determined by spectrophotometer at 280 nm. Concentration of the protein was performed using the Centricon centrifugal ultrafiltration device (Vivascience). The final concentration of His-CT2 was 50 mg/ml in 100 μ l.

7.2.4 Initial crystallization trials

The protein concentrations used for the initial crystal screens were 7, 15, 21, 34 and 50 mg/ml. Initial screens were set up using the sitting drop vapour diffusion method. The PACT, JCSG, PEG and Classics screen suites were used. Initial crystal screens were incubated in a temperature-controlled environment at 17 °C. Plates were checked after a few days and then every week after that. In all the crystallization attempts using different concentrations, the same unsatisfactory stage of crystal formation was achieved. Quasi crystals were produced in all of these trials (Figure 7.7). They were seen in the JCSG suite wells A11, B6, E9 and also in the Classic suite wells E11, F7, H9. Attempts were made to optimize the initial hits by reproducing the condition in 24-well plate format using the hanging drop vapour diffusion method, but better crystals were not grown.

7.2.5 Investigation of the CT2 oligomeric state by Electron Microscopy

To examine the CT2 oligomeric state by Electron Microscopy, diluted sample of His-CT2 (0.1 mg/ml) was negatively stained using 2% ammonium molybdate Stain, pH 7. Negatively stained electron images were recorded on a Philips CM100 transmission electron microscope, by a Gatan MultiScan 794 charge-coupled device camera (Ohi *et al*, 2004). The His-CT2 electron microscope image shows a variety of protein particle sizes. Particles of 65-75, 100 and 150-165 Å diameters were found suggesting dimers, trimers and tetramers, respectively (Figure 7.8). The heterogeneous state of His-CT2 may inhibit crystal growth. Pure, homogeneous protein is the most critical prerequisite for successful protein crystallization (Kundrot, 2004) and so further purification steps were required.

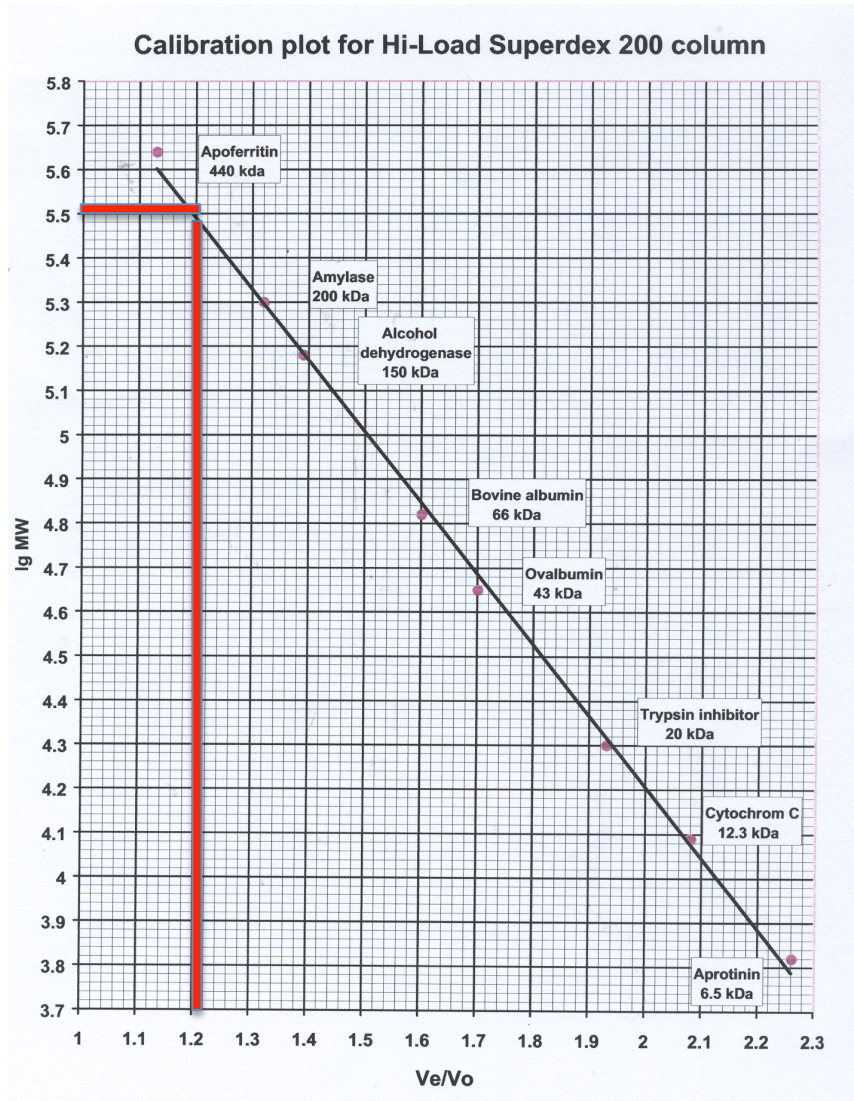


Figure 7.6: The calibration plot for Hi-Load superdex 200 column chart to calculate CT2 molecular weight: V_e (elution volume)/ V_o (column void volume)= 1.2, from the chart this gives a log MW of 5.5, therefore the estimated MW is ~ 350 KDa.



Figure 7.7: Quasi crystals of CT2 grew in 0.2 M ammonium phosphate, 0.1 M Tris pH 8.5 and 50% MPD.

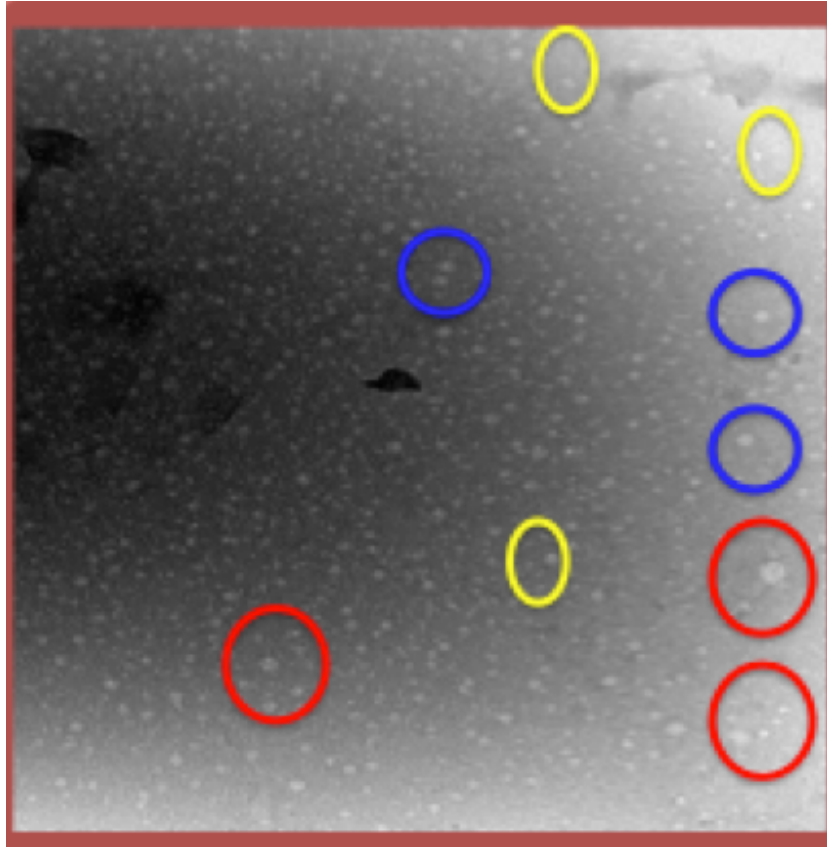


Figure 7.8: EM image for the oligomeric state of His-CT2. The red circle surround particles 100 Å diameter size which may be tetramers. The blue circles show particles of 65-80 Å diameter size suggesting trimers. The smallest particles are 50-60 Å diameter size and are surrounded by yellow rings, and suggest dimers.

7.2.6 Obtaining separated homogeneous fractions by using gel filtration

The previous gel filtration purification of His-CT2 and the Electron microscopy investigation show that CT2 tends to form oligomeric molecules, which might be the reason crystals do not grow. To change the heterogeneous state of CT2, two approaches have been taken: the first one was using gel filtration chromatography to obtain separated homogeneous fractions and the second was the introduction of a mutation in the predicted site of oligomerization. The gel filtration protocol was done as described before and the elution profile used to separate fractions (Figure 7.9). Two separated fractions were collected and reloaded on the gel filtration column. Initial crystallization trials have been prepared for each protein fraction. The two separated fractions behaved the same in gel filtration: two similar elution peaks were produced in the elution profiles. Also they produced quasi crystals, like those seen before.

7.2.7 Mutation in CT2 coiled-coil

CT2 has four hydrophobic residues in the CT2 predicted coiled coil (VAL846, ILE853, ILE860 and LEU867) which can be replaced with alanine to disrupt the coiled coil and prevent oligomerization. The mutant CT2 (CT2,4M) (the four hydrophobic residues in the coiled coil domain were replaced with alanine) was provided by the Sheffield University Medical School. The His-CT24M plasmid was transformed into competent cells (BL21-DE3-RIPL), and expressed using the same protocol as for the wild-type CT2 expression. The SDS gel shows that His-CT24M expressed as well as the wild-type. The soluble fraction was purified by Nickel affinity chromatography and gel filtration and then the pure protein was used to set down the initial crystallization trials, but once again no crystals were found (Figure 7.10). The oligomeric state of His-CT24M was also examined by Electron Microscopy. The EM image shows particles of 25-35 Å diameters (possibly monomers), but dimers also have been seen.

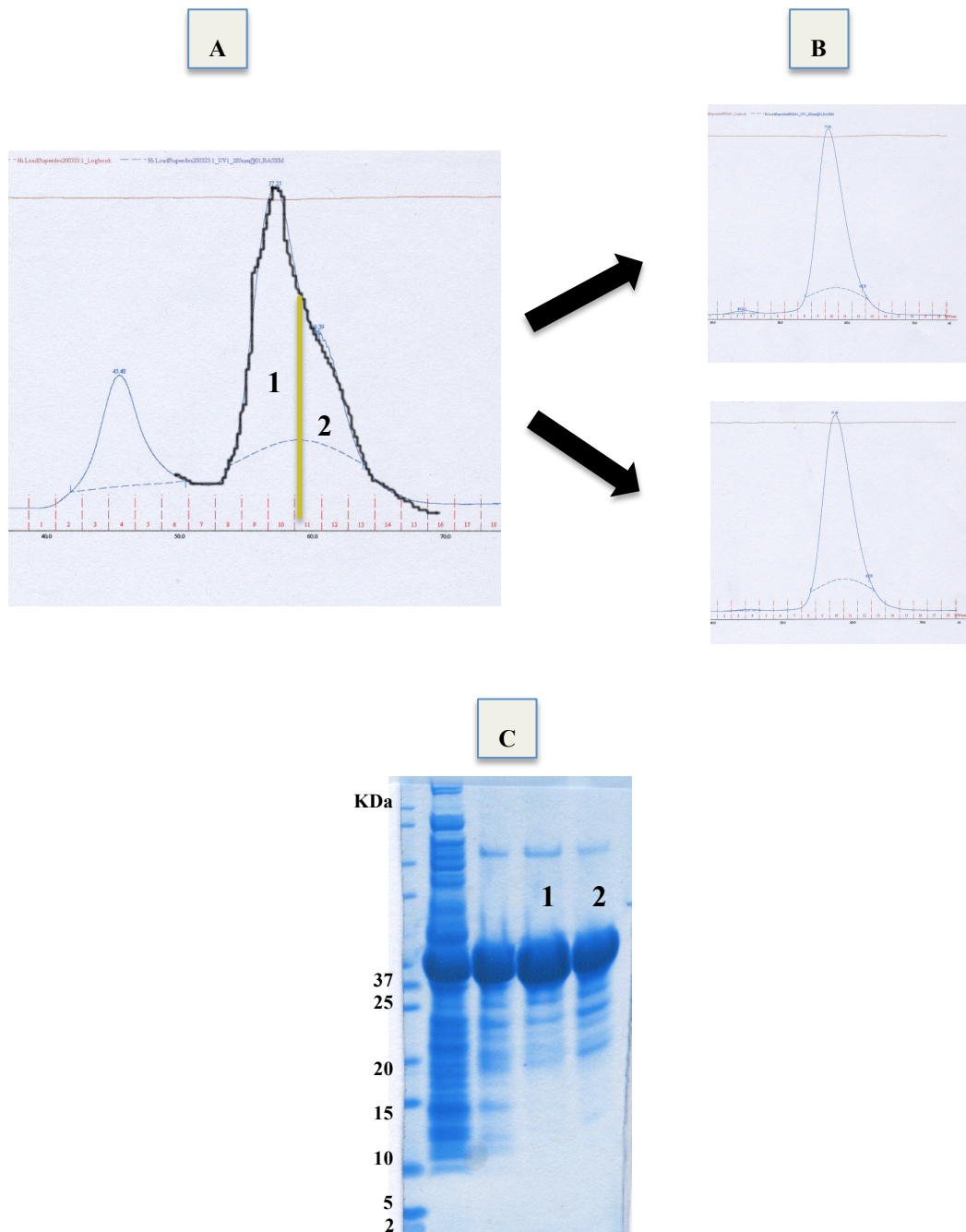


Figure 7.9: using gel filtration to achieve homogenous fractions of His-CT2. A: Elution profile divided into two peaks. B: Elution profiles of the reloaded separated peaks on gel filtration column. C: 12% SDS PAGE of fraction 1 and fraction 2 of 35 KDa each.

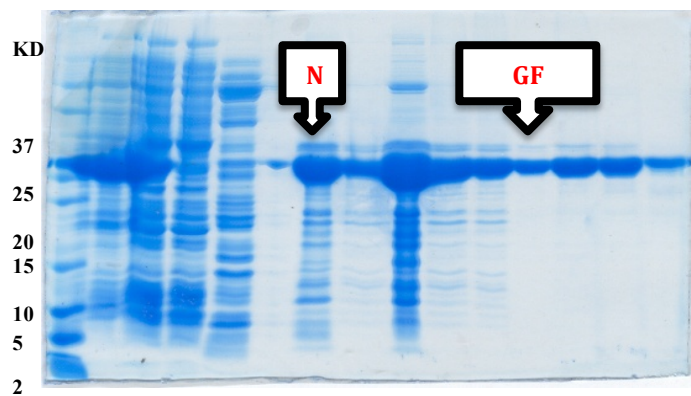
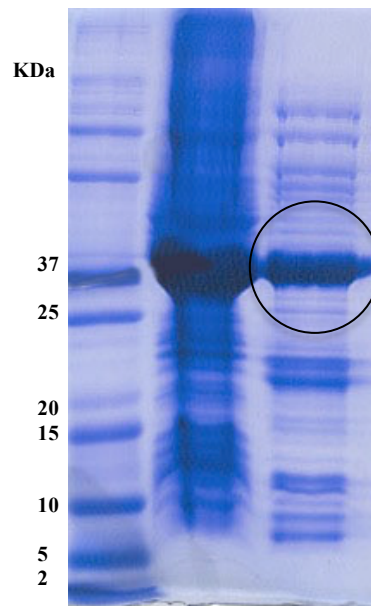


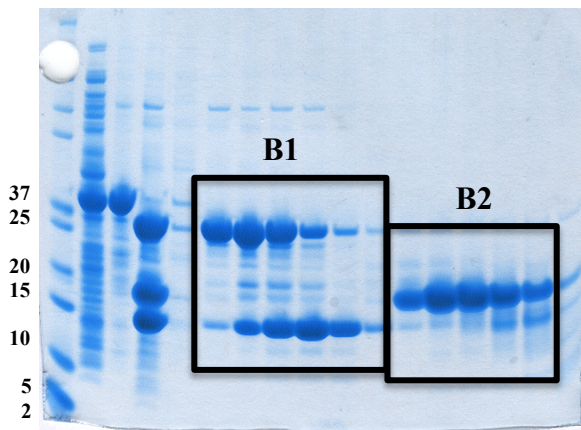
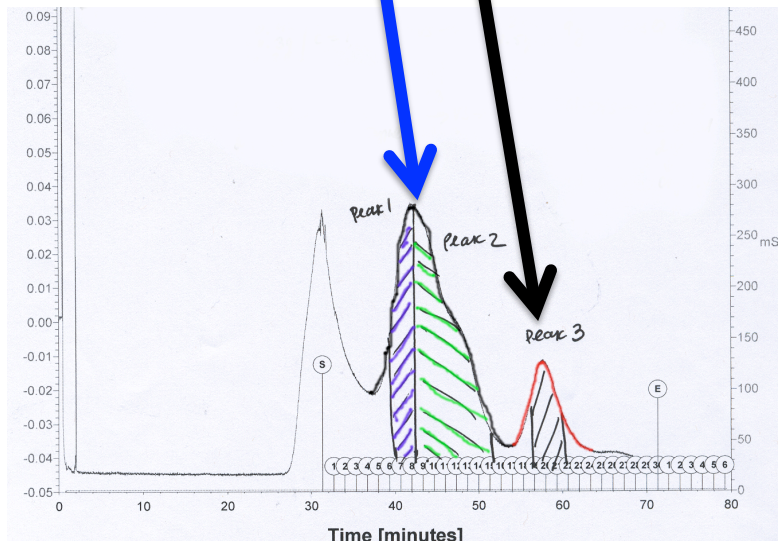
Figure 7.10: Mutation in CT2 coiled-coil. A: His-CT24M overexpression. B: His-CT24M Nickel column and gel filtration C: No crystals were found in crystallization trials

7.2.8 Removing the His-tag

The His-CT2 plasmid has a thrombin cleavage site between the His-tag and the CT2 protein (Figure 7.2). In attempts to improve the crystal formation, the 6 histidine tag was removed in the following way: wild type His-CT2 was expressed and purified on a Nickel column. The pure protein was dialyzed in thrombin digestion buffer (50% Hepes pH 7.4, 2.5 mg/ml CaCl₂, 250 mg NaCl) for 3 hours. The protein was then incubated with 50 units of thrombin overnight. The incubated mixture was loaded on a gel filtration column to elute the cleaved protein. The His-CT2 construct has by chance, in addition to the thrombin-cutting site (RGS) at the N-terminus, another cutting site in the following residues (ARG827, GLY828, SER829). Therefore thrombin cut at both cleavage sites and consequently CT2 was divided into two portions in addition to the His-tag removal. The N-terminal part is 146 residues (ILE680-ARG826) with a mass of 16.6 kDa and the C-terminal portion is 144 residues (ARG826-VAL970) with a molecular weight of 16.3 kDa. Although the two parts have approximately equal molecular weights, the gel filtration elution profile shows two different peaks (peak1 59.95 ml, peak2 44 ml) (Figure 7.11). The molecular weights of both portions were measured by using the gel filtration chart: the molecular weight of peak1 portion is ~ 35 kDa which suggests a dimer and the molecular weight of the peak2 portion is ~ 17 kDa suggesting a monomer. By applying both fragments to a SDS-PAGE gel, peak1 appeared at size ~ 35 kDa and peak2 at ~ 17 kDa (Figure 7.11). According to what is known about the CT2 domains, the monomer portion might be the N-terminal domain, which includes the EF-hand domain and the portion that ran as an oligomer might be the C-terminus, which contains the coiled-coil oligomerization domain. Although it might be expected that the dimer would run as a monomer in the denaturing SDS/PAGE condition, it has been observed that coiled coil proteins can remain as dimer on SDS/PAGE. However, these two separated proteins are not exciting targets as they have already been characterized structurally.

MGSSHHHHHSGLVPRGSH IINDTYSEVKSDLAQQKAEMELSD
 LIRKGYHKALVKLKLKNTVDDISESLRQGGGKLNFDLRODLKG
 KGHTDAEIEAIFTKYDQDGDQELTEHEHQMRDDLEKE EDLD
 LDHSSLPRPMSSRSFPRSLDDSEEDLDED SGHSSRRRGVSGVS
 YEEFQVLVRRVDRMEHSIGSIVSKIDAVIVKLEIMERAKLARREVL
 GRLLDGV AEDERLGRDSEIHREQMELVREELERWESDDAASQI
 SHGLGTPVGLNGQPRPRSSRPSSSQSTEGMEGAGGNGSSNVV

A



B

Figure 7.11: Removing the His-tag experiment. A: The N-terminal portion sequence represented as black letters and the C-terminal portion sequence as blue letters. The elution profile of the N-terminal portion (peak 44 ml) was indicated by a black arrow and the C-terminal portion (peak 59.95 ml) was indicated by a blue arrow. B: 12% SDS gel after gel filtration, where B1 represents the C-terminal portion (peak 59.95 ml) and B2 represents the N-terminal portion (peak 44 ml).

7.3 Conclusion and future work

During the course of this thesis we have successfully CT2 protein.. The C-terminal domain of Polycystin2 (CT2), which is predicted to consist of two domains, the EF-hand domain and the coil-coiled domain, is highly overexpressed in the *E. coli* and it can be purified successfully, producing a highly concentrated protein. Unfortunately this project has not reach the desired goal yet, although progress has been made, the protein was not able to generate crystals due to the heterogeneous status of the sample. The oligomeric state of CT2 was investigated using the gel filtration and electron microscopy. The chemical and physical characteristics of a protein are very important in its ability to crystallise. It might be useful to try the Fab mediated crystallization, which is a powerful technique to crystallise difficult proteins, particularly when one has access to a panel of monoclonal antibodies.

The CT1 the C-terminal domain of Polycystin 1 is a small protein, which might be explain the production of this protein as inclusion bodies even it has a binding fusion to increase the solubility. Several efforts have been done to refold CT1 using the protein refolding procedure that mentioned in thesis. Unfortunately we did not produce a significant amount of soluble protein. Future progress with this project may be achieved through the use of alternative vectors or another expression system, such as baculovirus, that may yield soluble CT1 protein.

Reference:

Abbink FC, Orendi JM, de Beaufort AJ (2001) Mother-to-child transmission of *Burkholderia pseudomallei*. *N Engl J Med* **344**: 1171-1172

Adelsberg J, Chamberlain S, Agati V (1997) Polycystic expression is temporally and spatially regulated during renal development. *Am J Physiol* **272**: 602-609

Aldhous P (2005) Tropical medicine: melioidosis? Never heard of it. *Nature* **434**: 692-693

Aravind L, Ponting CP (1997) The GAF domain: an evolutionary link between diverse phototransducing proteins. *Trends Biochem Sci* **22**: 458-459

Ashdown LR, Koehler JM (1990) Production of hemolysin and other extracellular enzymes by clinical isolates of *Pseudomonas pseudomallei*. *J Clin Microbiol* **28**: 2331-2334

AuCoin DP, Reed DE, Marlenee NL, Bowen RA, Thorkildson P, Judy BM, Torres AG, Kozel TR (2012) Polysaccharide specific monoclonal antibodies provide passive protection against intranasal challenge with *Burkholderia pseudomallei*. *PLoS ONE* **7**: e35386

Badger J, Sauder J, Adams J, Antonysamy S, Bain K, Bergseid M, Buchanan S, Buchanan M, Batiyenko Y, Christopher J, Emtage S, Eroshkina A, Feil I, Furlong E, Gajiwala K, Gao X, He D, Hendle J, Huber A, Hoda K, Kearins P, Kissinger C, Laubert B, Lewis H, Lin J, Loomis K, Lorimer D, Louie G, Maletic M, Marsh C, Miller I, Molinari J, Muller-Dieckmann H, Newman J, Noland B, Pagarigan B, Park F, Peat T, Post K, Radojicic S, Ramos A, Romero R, Rutter M, Sanderson W, Schwinn K, Tresser J, Winhoven J, Wright T, Wu L, Xu J, Harris T (2005) Structural analysis of a set of proteins resulting from a bacterial genomics project. *Proteins: Structure, Function, and Bioinformatics* **60**: 787-796

Barman P, Kaur R, Kumar K (2013) Clinically lesser known entity in India: A Report of two cases of Melioidosis. *Indian J Crit Care Med* **17**: 46-48

Bartlett JM, Stirling D (2003) A short history of the polymerase chain reaction. *Methods Mol Biol* **226**: 3-6

Behera B, Prasad Babu TL, Kamalesh A, Reddy G (2012) Ceftazidime resistance in *Burkholderia pseudomallei*: first report from India. *Asian Pac J Trop Med* **5**: 329-330

Blundell TL, Johnson LN (1976) *Protein Crystallography*, New York: Academic Press.

Bondi SK, Goldberg JB (2008) Strategies toward vaccines against *Burkholderia mallei* and *Burkholderia pseudomallei*. *Expert Rev Vaccines* **7**: 1357-1365

Bong SM, Chi YM. (2011) structure of fRMsR from *streptococcus pneumoniae*.

Bong SM, Kwak GH, Moon JH, Lee KS, Kim HS, Kim HY, Chi YM (2010) Structural and kinetic analysis of free methionine-R-sulfoxide reductase from *Staphylococcus aureus*: conformational changes during catalysis and implications for the catalytic and inhibitory mechanisms. *J Biol Chem* **285**: 25044-25052

Boschi-Muller S, Gand A, Branlant G (2008) The methionine sulfoxide reductases: Catalysis and substrate specificities. *Arch Biochem Biophys* **474**: 266-273

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* **72**: 248-254

Brändén C-I, Tooze J (2009) *Introduction to protein structure*, 2nd edn. New York, NY: Garland Pub.

Burtneck MN, Brett PJ, Harding SV, Ngugi SA, Ribot WJ, Chantratita N, Scorpio A, Milne TS, Dean RE, Fritz DL, Peacock SJ, Prior JL, Atkins TP, DeShazer D (2011) The Cluster 1 Type VI Secretion System Is a Major Virulence Determinant in *Burkholderia pseudomallei*. *Infect Immun* **79**: 1512-1525

Caldwell P, Luk DC, Weissbach H, Brot N (1978) Oxidation of the methionine residues of *Escherichia coli* ribosomal protein L12 decreases the protein's biological activity. *Proc Natl Acad Sci U S A* **75**: 5349-5352

Chan PH (1996) Role of oxidants in ischemic brain damage. *Stroke* **27**: 1124-1129

Chan YY, Chua KL (2005a) The *Burkholderia pseudomallei* BpeAB-OprB efflux pump: Expression and impact on quorum sensing and virulence. *J Bacteriol* **187**: 4707-4719

Chan YY, Chua KL (2005b) The *Burkholderia pseudomallei* BpeAB-OprB efflux pump: expression and impact on quorum sensing and virulence. *J Bacteriol* **187**: 4707-4719

Chantratita N, Wuthiekanun V, Boonbumrung K, TiyaWisutsri R, Vesaratchavest M, Limmathurotsakul D, Chierakul W, Wongratanacheewin S, Pukritiyakamee S, White NJ, Day NP, Peacock SJ (2007) Biological relevance of colony morphology and phenotypic switching by *Burkholderia pseudomallei*. *J Bacteriol* **189**: 807-817

Chaowagul W, White NJ, Dance DA, Wattanagoon Y, Naigowit P, Davis TM, Looareesuwan S, Pitakwatchara N (1989) Melioidosis: a major cause of community-acquired septicemia in northeastern Thailand. *J Infect Dis* **159**: 890-899

Chapin H, Caplan M (2010) The cell biology of polycystic kidney disease. *JCB* **191**: 701-710

Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**: 12-21

Chen YS, Chen SC, Kao CM, Chen YL (2003) Effects of soil pH, temperature and water content on the growth of *Burkholderia pseudomallei*. *Folia Microbiol (Praha)* **48**: 253-256

Cheng AC, Currie BJ (2005a) Melioidosis: epidemiology, pathophysiology, and management. *Clin Microbiol Rev* **18**: 383-416

Cheng AC, Currie BJ (2005b) Melioidosis: epidemiology, pathophysiology, and management. *Clin Microbiol Rev* **18**: 383-416

Chong CE, Lim BS, Nathan S, Mohamed R (2006) In silico analysis of *Burkholderia pseudomallei* genome sequence for potential drug targets. *In Silico Biol* **6**: 341-346

Chua KL, Chan YY, Gan YH (2003) Flagella are virulence determinants of *Burkholderia pseudomallei*. *Infect Immun* **71**: 1622-1629

Costerton JW, Stewart PS, Greenberg EP (1999) Bacterial biofilms: a common cause of persistent infections. *Science* **284**: 1318-1322

Currie BJ, Mayo M, ANSTEY NM, DONOHOE P, Haase G, KEMP DJ (2001) A CLUSTER OF MELIOIDOSIS CASES FROM AN ENDEMIC REGION IS CLONAL AND IS LINKED TO THE WATER SUPPLY USING MOLECULAR TYPING OF BURKHOLDERIA PSEUDOMALLEI ISOLATES. *Am J Trop Med Hyg* **65**: 177-179

Dance DA (1991a) Melioidosis: the tip of the iceberg? *Clin Microbiol Rev* **4**: 52-60

Dance DA (1991b) Melioidosis: the tip of the iceberg? *Clin Microbiol Rev* **4**: 52-60

Dance DA (2000a) Ecology of *Burkholderia pseudomallei* and the interactions between environmental *Burkholderia* spp. and human-animal hosts. *Acta Tropica* **74**: 159-168

Dance DA (2000b) Melioidosis as an emerging global problem. *Acta Tropica* **74**: 115-119

DeLano W, Lam J (2005) PyMOL: A communications tool for computational models. *Abstr Pap Am Chem Soc* **230**: 254-COMP

DeShazer D, Brett PJ, Burtnick MN, Woods DE (1999) Molecular characterization of genetic loci required for secretion of exoproducts in *Burkholderia pseudomallei*. *J Bacteriol* **181**: 4661-4664

Deshazer D, Brett PJ, Woods DE (1998) The type II O-antigenic polysaccharide moiety of *Burkholderia pseudomallei* lipopolysaccharide is required for serum resistance and virulence. *Molecular Microbiology* **30**: 1081-1100

Drenth J (1999) *Principles of Protein X-ray Crystallography*, 2nd edn. Berlin: Springer-Verlag.

El Hassouni M, Chambost JP, Expert D, Van Gijsegem F, Barras F (1999) The minimal gene set member *msrA*, encoding peptide methionine sulfoxide reductase, is a virulence determinant of the plant pathogen *Erwinia chrysanthemi*. *Proc Natl Acad Sci U S A* **96**: 887-892

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**: 2126-2132

Essex-Lopresti AE, Boddey JA, Thomas R, Smith MP, Hartley MG, Atkins T, Brown NF, Tsang CH, Peak IR, Hill J, Beacham IR, Titball RW (2005) A type IV pilin, PilA, Contributes To Adherence of *Burkholderia pseudomallei* and virulence in vivo. *Infect Immun* **73**: 1260-1264

Ezraty B, Aussel L, Barras F (2005) Methionine sulfoxide reductases in prokaryotes. *Biochim Biophys Acta* **1703**: 221-229

Finkel T (2000) Redox-dependent signal transduction. *FEBS Lett* **476**: 52-54

Fogazzi G (1998) The description of polycystic kidney by Domenico Gusmano Galeazzi
. *Nephrol Dial Transplant* **13**: 1039-1040

Foggensteiner L, Bevan A, Thomas R, Coleman N, Boulter C, Bradley J, Ibraghimova-Beskrovnaya O, Klinger K, Sandford R (2000) Cellular and subcellular distribution of polycystin-2, the protein product of the PKD2 gene. *J Am Soc Nephrol* **11**: 814-827

Francis SH, Lincoln TM, Corbin JD (1980) Characterization of a novel cGMP binding protein from rat lung. *J Biol Chem* **255**: 620-626

Gal D, Mayo M, Smith-Vaughan H, Dasari P, McKinnon M, Jacups SP, Urquhart AI, Hassell M, Currie BJ (2004) Contamination of hand wash detergent linked to occupationally acquired melioidosis. *Am J Trop Med Hyg* **71**: 360-362

Gamage AM, Shui G, Wenk MR, Chua KL (2011) N-Octanoylhomoserine lactone signalling mediated by the BpsI-BpsR quorum sensing system plays a major role in biofilm formation of *Burkholderia pseudomallei*. *Microbiology* **157**: 1176-1186

Gauthier YP, Thibault FM, Paucod JC, Vidal DR (2000) Protease production by *Burkholderia pseudomallei* and virulence in mice. *Acta Tropica* **74**: 215-220

Geng L, Segal Y, Peissel B, Deng N, Pei Y, Carone F, Rennke HG, Glucksmann-Kuis AM, Schneider MC, Ericsson M, Reeders ST, Zhou J (1996) Identification and localization of polycystin, the PKD1 gene product. *The Journal of clinical investigation* **98**: 2674-2682

Grimaud R, Ezraty B, Mitchell JK, Lafitte D, Briand C, Derrick PJ, Barras F (2001) Repair of oxidized proteins. Identification of a new methionine sulfoxide reductase. *J Biol Chem* **276**: 48915-48920

Gruetz A, Libiad M, Boschi-Muller S, Branlant G (2010) Structural and biochemical characterization of free methionine-R-sulfoxide reductase from *Neisseria meningitidis*. *J Biol Chem* **285**: 25033-25043

Hames D, Hooper N (2005) *Biochemistry*, Abingdon UK: Taylor & Francis Group.

Heikaus CC, Pandit J, Klevit RE (2009) Cyclic nucleotide binding GAF domains from phosphodiesterases: structural and mechanistic insights. *Structure* **17**: 1551-1557

Ho YJ, Burden LM, Hurley JH (2000) Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. *The EMBO Journal* **19**: 5288-5299

Holden MT, Titball RW, Peacock SJ, Cerdeno-Tarraga AM, Atkins T, Crossman LC, Pitt T, Churcher C, Mungall K, Bentley SD, Sebahia M, Thomson NR, Bason N, Beacham IR, Brooks K, Brown KA, Brown NF, Challis GL, Cherevach I, Chillingworth T, Cronin A, Crossett B, Davis P, DeShazer D, Feltwell T, Fraser A, Hance Z, Hauser H, Holroyd S, Jagels K, Keith KE, Maddison M, Moule S, Price C, Quail MA, Rabinowitsch E, Rutherford K, Sanders M, Simmonds M, Songsivilai S, Stevens K, Tumapa S, Vesaratchavest M, Whitehead S, Yeats C, Barrell BG, Oyston PC, Parkhill J (2004)

Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *Proc Natl Acad Sci U S A* **101**: 14240-14245

Holm L, Rosenstrom P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**: W545-549

Hoshi T, Heinemann S (2001) Regulation of cell function by methionine oxidation and reduction. *J Physiol* **531**: 1-11

Howard K, Inglis TJ (2003) The effect of free chlorine on *Burkholderia pseudomallei* in potable water. *Water Res* **37**: 4425-4432

Hughes J, Ward C, Peral B, Aspinwall R, Clark K, San Millan J, Gamble V, Harris P (1995) The polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell recognition domains. *Nat Genet* **10**: 151-160

Jones AL, Beveridge TJ, Woods DE (1996) Intracellular survival of *Burkholderia pseudomallei*. *Infect Immun* **64**: 782-790

Jones S, Thornton JM (1995) Protein-protein interactions: a review of protein dimer structures. *Progress in biophysics and molecular biology* **63**: 31-65

Kabsch W (1988) Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J Appl Crystallogr* **21**: 916-924

Kabsch W (2010a) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* **66**: 133-144

Kabsch W (2010b) Xds. *Acta Crystallogr D Biol Crystallogr* **66**: 125-132

Kaminski N (2000) Bioinformatics. A user's perspective. *Am J Respir Cell Mol Biol* **23**: 705-711

Kantardjieff K, Rupp B (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12**: 1865-1871

Kelley L, Sternberg M (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nature* **4**: 363-371

Kespichayawattana W, Rattanachetkul S, Wanun T, Utaisincharoen P, Sirisinha S (2000) *Burkholderia pseudomallei* induces cell fusion and actin-associated membrane protrusion: a possible mechanism for cell-to-cell spreading. *Infect Immun* **68**: 5377-5384

Kim HY, Gladyshev VN (2005a) Different catalytic mechanisms in mammalian selenocysteine- and cysteine-containing methionine-R-sulfoxide reductases. *PLoS Biol* **3**: e375

Kim HY, Gladyshev VN (2005b) Role of structural and functional elements of mouse methionine-S-sulfoxide reductase in its subcellular distribution. *Biochemistry* **44**: 8059-8067

Korbsrisate S, Tomaras AP, Damnin S, Ckumdee J, Srinon V, Lengwehasatit I, Vasil ML, Suparak S (2007) Characterization of two distinct phospholipase C enzymes from *Burkholderia pseudomallei*. *Microbiology* **153**: 1907-1915

Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology* **372**: 774-797

Kryukov GV, Kumar AR, Koc A, Sun Z, Gladyshev VN (2002) Selenoprotein R is a zinc-containing stereo-specific methionine sulfoxide reductase. *Proc Natl Acad Sci U S A* **99**: 4245-4250

Kumar A, Mayo M, Trunck LA, Cheng AC, Currie BJ, Schweizer HP (2008) Expression of resistance-nodulation-cell-division efflux pumps in commonly used *Burkholderia pseudomallei* strains and clinical

isolates from northern Australia. *Trans R Soc Trop Med Hyg* **102**: 45-51

Kuschel L, Hansel A, Schonherr R, Weissbach H, Brot N, Hoshi T, Heinemann SH (1999) Molecular cloning and functional expression of a human peptide methionine sulfoxide reductase (hMsrA). *FEBS Lett* **456**: 17-21

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685

Laskowski RA (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26**: 283-291

Lee BC, Dikiy A, Kim HY, Gladyshev VN (2009) Functions and evolution of selenoprotein methionine sulfoxide reductases. *Biochim Biophys Acta* **1790**: 1471-1477

Leslie AGW (1994) MOSFLM. Joint CCP4 + ESF-EAMCB. *Newsletter on Protein Crystallography* **26**

Lim WK, Gurdeep GS, Norain K (2001) Melioidosis of the head and neck. *Med J Malaysia* **56**: 471-477

Lin Z, Johnson LC, Weissbach H, Brot N, Lively MO, Lowther WT (2007) Free methionine-(R)-sulfoxide reductase from *Escherichia coli* reveals a new GAF domain function. *Proc Natl Acad Sci U S A* **104**: 9597-9602

Ling JLL, Nathan S, Hin LK, Mohamed R (2001) Purification and Characterisation of a *Burkholderia pseudomallei* Protease Expressed in Recombinant *E. coli*. *Journal of Biochemistry and Molecular Biology* **34**: 509-516

Lowther WT, Brot N, Weissbach H, Honck JF, Matthews BW (2000) Thiol–disulfide exchange is involved in the catalytic mechanism of peptide methionine sulfoxide reductase. *Proc Natl Acad Sci U S A* **97**: 6463–6468

Lu W, Peissel B, Babakhanlou H, Pavlova A, Geng L, Fan X, Larson C, Brent G, Zhou J (1997) Perinatal lethality with kidney and pancreas defects in mice with a targeted *pkd1* mutation. *Nat Genet* **17**: 179-181

Luo Y, Vassilev P, Xiaogang L, Yoshifumi K, Zhou J (2003) Native Polycystin 2 Functions as a Plasma Membrane Ca²⁺-Permeable Cation Channel in Renal Epitheli. *MOLECULAR AND CELLULAR BIOLOGY* **23** 2600-2607

Martinez S, Wu A, Glavas N, Tang X, Turley S, Hol W (2002a) The two GAF domains in phosphodiesterase 2A have distinct roles in dimerization and in cGMP binding. *PNAS* **99**: 13260-13265

Martinez SE, Beavo JA, Hol WGJ (2002b) GAF Domains: Two-Billion-Year-Old Molecular Switches that Bind Cyclic Nucleotides. *Molecular interventions* **2**: 317-323

Martinez SE, Bruder S, Schultz A, Zheng N, Schultz JE, Beavo JA, Linder JU (2005) Crystal structure of the tandem GAF domains from a cyanobacterial adenylyl cyclase: modes of ligand binding and dimerization. *Proc Natl Acad Sci U S A* **102**: 3082-3087

Matthews BW (1968) Solvent content of protein crystals. *Journal of molecular biology* **33**: 491-497

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**: 658-674

Meyer Y, Buchanan BB, Vignols F, Reichheld JP (2009) Thioredoxins and glutaredoxins: unifying elements in redox biology. *Annual review of genetics* **43**: 335-367

Mima T, Schweizer HP (2010) The BpeAB-OprB efflux pump of *Burkholderia pseudomallei* 1026b does not play a role in quorum sensing, virulence factor production, or extrusion of aminoglycosides but is a broad-spectrum drug efflux system. *Antimicrob Agents Chemother* **54**: 3113-3120

Mochizuki T, Wu G, Hayashi T, Xenophontos S, Veldhuisen B, Saris J, Reynolds D, Cai Y, Gabow P, Pierides A, Kimberling W, Breuning M, Deltas C, Peters D, Somlo S (1996) PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. *Science* **272**: 1339-1342

Moore RA, DeShazer D, Reckseidler S, Weissman A, Woods DE (1999) Efflux-mediated aminoglycoside and macrolide resistance in *Burkholderia pseudomallei*. *Antimicrob Agents Chemother* **43**: 465-470

Moskovitz J, Poston JM, Berlett BS, Nosworthy NJ, Szczepanowski R, Stadtman ER (2000) Identification and characterization of a putative active site for peptide methionine sulfoxide reductase (MsrA) and its substrate stereospecificity. *J Biol Chem* **275**: 14167-14172

Moskovitz J, Weissbach H, Brot N (1996) Cloning the expression of a mammalian gene involved in the reduction of methionine sulfoxide residues in proteins. *Proc Natl Acad Sci U S A* **93**: 2095-2099

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**: 240-255

Nair U, Bartsch H, Nair J (2007) Lipid peroxidation-induced DNA damage in cancer-prone inflammatory diseases: a review of published adduct types and levels in humans. *Free Radic Biol Med* **43**: 1109-1120

Nieves W, Asakrah S, Qazi O, Brown KA, Kurtz J, Aucoin DP, McLachlan JB, Roy CJ, Morici LA (2011) A naturally derived outer-membrane vesicle vaccine protects against lethal pulmonary *Burkholderia pseudomallei* infection. *Vaccine* **29**: 8381-8389

Nikaido H, Vaara M (1985) Molecular basis of bacterial outer membrane permeability. *Microbiol Rev* **49**: 1-32

O'Riordan K, Lee JC (2004) *Staphylococcus aureus* capsular polysaccharides. *Clin Microbiol Rev* **17**: 218-234

Ohi M, Li Y, Cheng Y, Walz T (2004) Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. *Biol Proced Online* **6**: 23-34

Ong A, Ward C, Butler R, Biddolph S, Bowker C, Torra R, Pei Y, Harris P (1999) Coordinate expression of the autosomal dominant polycystic kidney disease protein, polycystin-2 and polycystin-1, in normal and cystic tissue. *Am J Pathol* **154**: 1721-1729

Peacock SJ, Limmathurotsakul D, Lubell Y, Koh GC, White LJ, Day NP, Titball RW (2012) Melioidosis vaccines: a systematic review and appraisal of the potential to exploit biodefense vaccines for public health purposes. *PLoS Negl Trop Dis* **6**: e1488

Pei J, Kim BH, Grishin NV (2008) PROMALS3D: a tool for multiple sequence and structure alignment. *Nucleic Acids Research* **36**: 2295-2300

Perry MB, MacLean LL, Schollaardt T, Bryan LE, Ho M (1995) Structural characterization of the lipopolysaccharide O antigens of *Burkholderia pseudomallei*. *Infect Immun* **63**: 3348-3352

Petri E, Čelića A, Kennedy S, Ehrlich B, Boggona T, Hodsdon M (2010) Structure of the EF-hand domain of polycystin-2 suggests a

mechanism for Ca²⁺-dependent regulation of polycystin-2 channel activity. *PNAS Proceeding of the National Academy of Sciences of the United States of America* **107**: 9176-9181

Pumpuang A, Chantratita N, Wikraiphath C, Saiprom N, Day NP, Peacock SJ, Wuthiekanun V (2011) Survival of *Burkholderia pseudomallei* in distilled water for 16 years. *Trans R Soc Trop Med Hyg* **105**: 598-600

Puthuchery SD, Vadivelu J, Ce-Cile C, Kum-Thong W, Ismail G (1996) Short report: Electron microscopic demonstration of extracellular structure of *Burkholderia pseudomallei*. *Am J Trop Med Hyg* **54**: 313-314

Qian F, Germino F, Cai Y (1997) PKD1 interacts with PKD2 through a probable coiled-coil domain. *Nat Genet* **16**: 179-183

Reckseidler SL, Deshazer D, Sokol PA, Woods DE (2001) Detection of Bacterial Virulence Genes by Subtractive Hybridization: Identification of Capsular Polysaccharide of *Burkholderia pseudomallei* as a Major Virulence Determinant. *Infect Immun* **69**: 34

Reckseidler-Zenteno SL, DeVinney R, Woods DE (2005) The capsular polysaccharide of *Burkholderia pseudomallei* contributes to survival in serum by reducing complement factor C3b deposition. *Infect Immun* **73**: 1106-1115

Reeders S, Keith T, Green P, Germino G, Barton N, Lehmann O, Brown V, Phipps P, Morgan J, Bear J (1988) regional localization of the autosomal dominant polycystic kidney disease locus. *Genomics* **3**: 150-155

Rhodes G (2006) *Crystallography made crystal clear : a guide for users of macromolecular models*, 3rd edn. Amsterdam ; Boston: Elsevier/Academic Press.

Rotz LD, Khan AS, Lillibridge SR, Ostroff SM, Hughes JM (2002) Public health assessment of potential biological terrorism agents. *Emerg Infect Dis* **8**: 225-230

Rupp B (2010) *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, New York: Garland Science.

Sambrook JW, Russell DW (2006) *The condensed protocols from molecular cloning: a Laboratory Manual*, New York, USA: Cold Spring Harbor Press.

Sandowski Y, Raver N, Gussakovsky EE, Shochat S, Dym O, Livnah O, Rubinstein M, Krishna R, Gertler A (2002) Subcloning, expression, purification, and characterization of recombinant human leptin-binding domain. *J Biol Chem* **277**: 46304-46309

Sawasdidoln C, Taweechaisupapong S, Sermswan RW, Tattawasart U, Tungpradabkul S, Wongratanacheewin S (2010) Growing *Burkholderia pseudomallei* in biofilm stimulating conditions significantly induces antimicrobial resistance. *PLoS ONE* **5**

Shafer WM, Rest RF (1989) Interactions of gonococci with phagocytic cells. *Annu Rev Microbiol* **43**: 121-145

Shalom G, Shaw JG, Thomas MS (2007) In vivo expression technology identifies a type VI secretion system locus in *Burkholderia pseudomallei* that is induced upon invasion of macrophages. *Microbiology* **153**: 2689-2699

Singh VK, Moskovitz J (2003) Multiple methionine sulfoxide reductase genes in *Staphylococcus aureus*: expression of activity and roles in tolerance of oxidative stress. *Microbiology* **149**: 2739-2747

Stadtman ER (1992) Protein oxidation and aging. *Science* **257**: 1220-1224

Stadtman ER, Berlett BS (1998) Reactive oxygen-mediated protein oxidation in aging and disease. *Drug Metab Rev* **30**: 225-243

Stawicki S, Lombardo G (2008) ABSITE CORNER Polycystic kidney disease. *OPUS 12 Scientist* **2**: 17-20

Stevens MP, Wood MW, Taylor LA, Monaghan P, Hawes P, Jones PW, Wallis TS, Galyov EE (2002) An Inv/Mxi-Spa-like type III protein secretion system in *Burkholderia pseudomallei* modulates intracellular behaviour of the pathogen. *Molecular Microbiology* **46**: 649-659

Stone R (2007) Infectious disease. Racing to defuse a bacterial time bomb. *Science* **317**: 1022-1024

Suparak S, Kespichayawattana W, Haque A, Easton A, Damnin S, Lertmemongkolchai G, Bancroft GJ, Korbsrisate S (2005) Multinucleated Giant Cell Formation and Apoptosis in Infected Host Cells Is Mediated by *Burkholderia pseudomallei* Type III Secretion Protein BipB. *J Bacteriol* **187**: 6556-6560

Suputtamongkol Y, Chaowagul W, Chetchotisakd P, Lertpatanasuwun N, Intaranongpai S, Ruchutrakool T, Budhsarawong D, Mootsikapun P, Wuthiekanun V, Teerawatasook N, Lulitanond A (1999) Risk factors for melioidosis and bacteremic melioidosis. *Clin Infect Dis* **29**: 408-413

Taweechaisupapong S, Kaewpa C, Arunyanart C, Kanla P, Homchampa P, Sirisinha S, Prongvitaya T, Wongratanacheewin S (2005) Virulence of *Burkholderia pseudomallei* does not correlate with biofilm formation. *Microbial Pathogenesis* **39**: 77-85

Thibault FM, Hernandez E, Vidal DR, Girardet M, Cavallo JD (2004) Antibiotic susceptibility of 65 isolates of *Burkholderia pseudomallei* and *Burkholderia mallei* to 35 antimicrobial agents. *J Antimicrob Chemother* **54**: 1134-1138

TSIOKAS L, KIM E, ARNOULD T, SUKHATME VP, WALZ G (1997) Homo- and heterodimeric interactions between the gene products of PKD1 and PKD2. *Proc Natl Acad Sci U S A* **94**: 6965-6970

Tu Z, He G, Li K, Chen M, Chang J, Chen L, Yao Q, Liu D, Ye H, Shi J, Wu X (2005) An improved system for competent cell preparation and high efficiency plasmid transformation using different *Escherichia coli* strains. *Electronic Journal of Biotechnology* **8**

Ulrich RL, Deshazer D, Brueggemann EE, Hines HB, Oyston PC, Jeddeloh JA (2004) Role of quorum sensing in the pathogenicity of *Burkholderia pseudomallei*. *J Med Microbiol* **53**: 1053-1064

Vorachit M, Lam K, Jayanetra P, Costerton JW (1995) Electron microscopy study of the mode of growth of *Pseudomonas pseudomallei* in vitro and in vivo. *J Trop Med Hyg* **98**: 379-391

Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein engineering* **8**: 127-134

Warawa J, Woods DE (2005) Type III secretion system cluster 3 is required for maximal virulence of *Burkholderia pseudomallei* in a hamster infection model. *FEMS Microbiology Letters* **242**: 101-108

Ward C, Turley H, Ong A, Comley M, Biddolph S, Chetty R, Ratcliffe P, Gatter K, P H (1996) Polycystin, the polycystin kidney disease 1 protein, is expressed by epithelial cell in fetal, and adult and polycystic kidney. *Proc Natl Acad Sci U S A* **93**: 1524-1528

Weissbach H, Resnick L, Brot N (2005) Methionine sulfoxide reductases: history and cellular role in protecting against oxidative damage. *Biochim Biophys Acta* **1703**: 203-212

White NJ (2003a) Melioidosis. *Lancet* **361**: 1715-1722

- White NJ (2003b) Melioidosis. *Lancet* **361**: 1715-1722
- White NJ, Dance DA, Chaowagul W, Wattanagoon Y, Wuthiekanun V, Pitakwatchara N (1989) Halving of mortality of severe melioidosis by ceftazidime. *Lancet* **2**: 697-701
- Wiersinga WJ, Poll T, White NJ, Day NP, Peacock SJ (2006) Melioidosis: insights into the pathogenicity of *Burkholderia pseudomallei*. *NATURE REVIEWS | MICROBIOLOGY* **4**: 272-282
- Wiersinga WJ, Poll T, White NJ, Day NP, Peacock SJ (2007) Melioidosis: Insights into the pathogenicity of *Burkholderia pseudomallei*. *NATURE REVIEWS | MICROBIOLOGY* **4**: 272-282
- Wilson P (2001) Polycystin: New Aspect Of Structure, Function, and Regulation. *Am Soc Nephrol* **12**: 834-845
- Wilson P (2004) Mechanism of Disease Polycystic Kidney Disease. . *N Engl J Med* **350**: 151-164
- Winrow VR, Winyard PG, Morris CJ, Blake DR (1993) Free radicals in inflammation: second messengers and mediators of tissue destruction. *Br Med Bull* **49**: 506-522
- Winter G (2010) xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* **43**
- Wongtrakoongate P, Mongkoldhumrongkul N, Chaijan S, Kamchonwongpaisanc S, Tungpradabkul S (2007) Comparative proteomic profiles and the potential markers between *Burkholderia pseudomallei* and *Burkholderia thailandensis*. *Molecular and Cellular Probes* **21**: 81-91
- Yang HM, Chaowagul W, Sokol PA (1991) Siderophore production by *Pseudomonas pseudomallei*. *Infect Immun* **59**: 776-780

Yee KC, Lee MK, Chua CT, Puthucheary SD (1988) Melioidosis, the great mimicker: a report of 10 cases from Malaysia. *J Trop Med Hyg* **91**: 249-254

Yu BP, Kang CM, Han JS, Kim DS (1998) Can antioxidant supplementation slow the aging process? *Biofactors* **7**: 93—101

Yu Y, Ulbrich MH, Li MH, Buraei Z, Chen XZ, Ong AC, Tong L, Isacoff EY, Yang J (2009) Structural and molecular basis of the assembly of the TRPP2/PKD1 complex. *Proc Natl Acad Sci U S A* **106**: 11558-11563

Zoraghi R, Corbin JD, Francis SH (2004) Properties and functions of GAF domains in cyclic nucleotide phosphodiesterases and other proteins. *Mol Pharmacol* **65**: 267-278

ee))